

**U. S. Army Research Office**

**Report No. ~~93-T~~ 93-2**

**June 1993**

**PROCEEDINGS OF THE THIRTY-EIGHTH CONFERENCE  
ON THE DESIGN OF EXPERIMENTS**

**Sponsored by the Army Mathematics Steering Committee**

**HOST**

**The Arroyo Center of the RAND Corporation  
Santa Monica, California**

**28-30 October 1992**

**Approved for public release; distribution unlimited.  
The findings in this report are not to be construed  
as an official Department of the Army position, un-  
less so designated by other authorized documents.**

**U.S. Army Research Office  
P. O. Box 12211  
Research Triangle Park, North Carolina**

**20010606 081**

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

## REPORT DOCUMENTATION PAGE

Form Approved  
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified			1b. RESTRICTIVE MARKINGS	
2a. SECURITY CLASSIFICATION AUTHORITY			3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release: distribution unlimited.	
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE			5. MONITORING ORGANIZATION REPORT NUMBER(S)	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) ARO REPORT 93-2			7a. NAME OF MONITORING ORGANIZATION	
6a. NAME OF PERFORMING ORGANIZATION Army Research Office		6b. OFFICE SYMBOL (if applicable) AMXRO-MCS	7b. ADDRESS (City, State, and ZIP Code)	
6c. ADDRESS (City, State, and ZIP Code) P. O. Box 12211 Research Triangle Park, NC 27709			9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER	
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AMSC		8b. OFFICE SYMBOL (if applicable)	10. SOURCE OF FUNDING NUMBERS	
8c. ADDRESS (City, State, and ZIP Code)			PROGRAM ELEMENT NO.	PROJECT NO.
			TASK NO.	WORK UNIT ACCESSION NO.
11. TITLE (Include Security Classification) Proceedings of the Thirty-Eighth Conference on the Design of Experiments in Army Research, Development and Testing				
12. PERSONAL AUTHOR(S)				
13a. TYPE OF REPORT Technical		13b. TIME COVERED FROM Jan 93 TO Feb 94	14. DATE OF REPORT (Year, Month, Day) 1993 June	15. PAGE COUNT 354
16. SUPPLEMENTARY NOTATION				
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)	
FIELD	GROUP	SUB-GROUP		
19. ABSTRACT (Continue on reverse if necessary and identify by block number)				
This is a technical report of the Thirty-Eighth Conference on the Design of Experiments in Army Research, Development and Testing. It contains most of the papers presented at this meeting. These articles treat various Army Statistical and design problems.				
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT <input checked="" type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS			21. ABSTRACT SECURITY CLASSIFICATION	
22a. NAME OF RESPONSIBLE INDIVIDUAL Dr. Francis G. Dressel			22b. TELEPHONE (Include Area Code) 919-549-4319	22c. OFFICE SYMBOL AMXRO-MCS

## FOREWORD

The host for the Thirty-Eighth Conference on the Design of Experiments in Army Research, Development and Testing (DOE) was the RAND Corporation. It was held on 28-30 October 1992 at the Arroyo Center of RAND in Santa Monica, California. In the host invitational letter Dr. Lynn E. Davis, Director of the Arroyo Center, stated that Ms. Sharon Koga and Major Kelvin Beam would be in charge of local arrangements. To conduct a conference of this size in a classified facility is a formidable task in itself, but they did it flawlessly down to the last detail. The participation of the excellent statistics group at RAND contributed to the success of the conference.

Members of the Program Committee for this conference were pleased to obtain the following invited speakers to talk on topics of current interest to Army personnel.

<u>Speaker and Affiliation</u>	<u>Title of Address</u>
Professor Donald Gaver Naval Postgraduate School	Simulation and Modeling
Professor David W. Scott Rice University	Visualization of Response Surfaces in Several Variables
Professor Nozer D. Singpurwalla George Washington University	Statistical Methods in Software Engineering
Dr. James J. Rissanen IBM Research Center, Almaden	Information Theory and Statistics
Professor L. Mark Berliner Ohio State University	Chaotic Systems and Statistics

In addition to the invited addresses, there were fourteen contributed papers, four clinical papers, three papers in a special session, and one paper in a poster session. Most of these informative talks covered areas associated with statistical design and analysis of experiments.

Dr. Malcolm S. Taylor of the Army Research Laboratory at Aberdeen Proving Ground, Maryland was the recipient of the Eleventh U.S. Army Wilks Award for contributions to statistical methodologies in Army Research, Development and Testing. In as much as the major part of his work has been to help military people with real problems that had to be solved, he has constantly and effectively employed a widely ranging arsenal of mathematical specialties to assist his clients. His published research includes work in computer science, experimental design, extreme value theory, nonparametric testing,

resampling theory, survival theory, fuzzy set theory, non-Newtonian flow, nonlinear programming, linear models, information theory, vulnerability theory, artificial intelligence, and control theory. Over the years, Malcolm Taylor has worn many hats, worked effectively on many problems, always in the best traditions of U.S. Military Science.

On 26-27 October 1992, two days before the start of the Design of Experiments Conference, a two day tutorial entitled "Statistics for Spacial Data" was held. Its speaker was Professor Noel Cressie of Iowa State University, Ames, Iowa. The main purpose of these seminars is to develop, in Army scientists, an interest in and an appreciation for the statistical methods that are needed to analyze experimental data.

The sponsor of these conferences is the Army Mathematics Steering Committee. Members of this committee would like to thank the RAND Corporation for hosting this conference and Dr. Lynn Davis for serving as Chairperson on local arrangements.

#### **PROGRAM COMMITTEE**

Carl Bates  
Eugene Dutoit  
Douglas Tang  
Barry Bodt

Robert Burge  
Malcolm Taylor  
Henry Tingey  
Gerald Andersen

Francis Dressel  
Carl Russell  
Jerry Thomas  
Jock Grynovicki



## TABLE OF CONTENTS\*

<u>Title</u>	<u>Page</u>
Foreword .....	iii
Table of Contents .....	v
Agenda .....	ix
 Modeling and Simulation in the Military: Statistical issues and Opportunities Donald P. Gaver .....	   1
 A Comparative Study of Boresight Devices for Tank Cannon David W. Webb and MAJ Bruce J. Held .....	  31
 The Rank Transformation in Balanced incomplete Block Designs W. J. Conover .....	  41
 The Application of Meta-Analysis to Army Issues Carl B. Bates and Franklin E. Womack .....	  53
 Total Time on Test Function Orthogonal Components and Tests of Exponentiality W. D. Kaigh and Alexander K. White .....	  65
 Determination of the Economic Acceptable Quality Level (EAQL) J. Steve Caruso .....	  87
 Some Problems of Estimation and Testing in Multivariate Statistical Process Control Martin Lawera and James R. Thompson .....	  99
 Sampling Problems Pertaining to the Number of Replications for Stochastic Simulation Models W. E. Baker, D W. Webb and L. D. Losie .....	  127

\*This Table of Contents contains only the papers that are published in this manual.

<u>Title</u>	<u>Page</u>
Formalizing the Determination of Spall Cone Angle Barry A. Bodt . . . . .	131
Models for Assessing the Reliability of Computer Software N. D. Singpurwalla and S. P. Wilson . . . . .	139
Statistical Methods Applied to Vocational Counseling Data Obtained From Army Veterans Gene Dutoit and John Mobley . . . . .	163
The MDL Principle - A Tutorial J. Rissanen . . . . .	175
Approximate One-Sided Tolerance Limits for a Mixed Model With a Nested Random Effect Mark G. Vangel . . . . .	203
Determination of Camouflage Effectiveness of Small Area Camouflage Covers (SACC) By Ground Observers Using the Method of Limit G. Anitole, R. Johnson and C. J. Neubert . . . . .	217
An Exploratory Algorithm for the Estimation of Mode Location and Numerosity in Multidimensional Data Mare N. Elliott and James R. Thompson . . . . .	229
Sample-Weighted Average of Vectors Aivars Celmins . . . . .	245
Jump Characterization Test Charles E. Heatwole . . . . .	259
Easy-to-Apply Results for Establishing Convergence of Markov Chains in Bayesian Analysis K. B. Athreya, H. Doss and J. Sethuraman . . . . .	263
Assessment of Helicopter Component Statistical Reliability Computations Donald Neal and William Matthews . . . . .	271
Wavelets and Nonparametric Function Estimation: A Function Analytic Approach Edward J. Wegman . . . . .	293

<u>Title</u>	<u>Page</u>
A Parallelized, Simulation Based Algorithm for Parameter Estimation Martin Lawera and James R. Thompson . . . . .	321
Simulation Based Estimation for Birth and Death Processes K. B. Ensor, E. Bridges and M. Lawera . . . . .	343
Attendance List . . . . .	353

**The Thirty Eighth Conference on the Design of Experiments  
in Army Research, Development and Testing**

**RAND, Santa Monica, California**

**Wednesday, 28 October 1992**

0800 - 0915	<b>REGISTRATION</b>	Main Conference Room Lobby
0800 - 0915	<b>CONTINENTAL BREAKFAST</b>	Common Room
0915 - 0930	<b>CALL TO ORDER</b> Main Conference Room	MAJ Kevin M. Beam US Army TRADOC, RAND
	<b>OPENING REMARKS</b> Main Conference Room	Lynn E. Davis Vice President, Army Research Division Director, Arroyo Center, RAND
0930 - 1200	<b>GENERAL SESSION I</b> Main Conference Room	Chairperson: MAJ Kevin M. Beam US Army TRADOC, RAND
0930-1030	<b>KEYNOTE ADDRESS</b> <b>SIMULATION AND MODELING</b>	Donald P. Gaver Naval Postgraduate School
1030-1100	<b>BREAK</b>	Common Room
1100-1200	<b>VISUALIZATION OF RESPONSE</b> <b>SURFACES IN SEVERAL VARIABLES</b>	David W. Scott Rice University
1200 - 1330	<b>LUNCH</b>	
1330 - 1500	<b>CONTRIBUTED SESSION I</b> Main Conference Room	Chairperson: Douglas B. Tang Walter Reed Army Institute of Research
	<b>A COMPARATIVE STUDY OF</b> <b>BORESIGHT DEVICES FOR TANK</b> <b>CANNON</b>	David W. Webb and MAJ Bruce J. Held US Army Research Laboratory
	<b>THE RANK TRANSFORMATION IN</b> <b>BALANCED INCOMPLETE BLOCK</b> <b>DESIGNS</b>	W. J. Conover Texas Tech University
	<b>THE APPLICATION OF META-</b> <b>ANALYSIS TO ARMY ISSUES</b>	Carl B. Bates and Franklin E. Womack US Army Concepts Analysis Agency
1500 - 1530	<b>BREAK</b>	Common Room

**Wednesday, 28 October 1992**

**1530 - 1700 CONTRIBUTED SESSION II**

Main Conference Room  
(note dual sessions)

Chairperson: William Jackson  
US ARMY TACOM

TOTAL TIME ON TEST FUNCTION  
ORTHOGONAL COMPONENTS AND  
TESTS OF EXPONENTIALITY

W. D. Kaigh  
University of Texas at El Paso

DETERMINATION OF THE ECONOMIC J. Steve Caruso  
ACCEPTABLE QUALITY LEVEL (EAQL) US Army Management Engineering College

SOME PROBLEMS OF ESTIMATION  
AND TESTING IN MULTIVARIATE  
STATISTICAL PROCESS CONTROL

Martin Lawera and James R. Thompson  
Rice University

**1530 - 1700 CLINICAL SESSION I**

Common Room  
(note dual sessions)

Chairperson: Terence M. Cronin  
US ARMY CECOM

Discussants: Russell R. Barton  
Penn State University  
Donald P. Gaver  
Naval Postgraduate School  
James S. Hodges  
RAND  
David W. Scott  
Rice University  
Nozer D. Singpurwalla  
George Washington University

SAMPLING PROBLEMS PERTAINING  
TO THE NUMBER OF REPLICATIONS  
FOR STOCHASTIC SIMULATION  
MODELS

William E. Baker and David W. Webb  
US Army Research Laboratory

FORMALIZING THE DETERMINATION  
OF SPALL CONE ANGLE

Barry A. Bodt  
US Army Research Laboratory

**1830 - WILKS AWARD BANQUET**

Toppers, Radison Huntley Hotel  
1111 2nd Street (corner of Wilshire and 2nd)

1830-1930 CASH BAR

1930- DINNER

# **The Thirty Eighth Conference on the Design of Experiments in Army Research, Development and Testing**

**RAND, Santa Monica, California**

**Thursday, 29 October 1992**

0730 - 0800	CONTINENTAL BREAKFAST	Common Room
0800 - 0900	GENERAL SESSION II Main Conference Room	Chairperson: Ann E. M. Brodeen US Army Research Laboratory
0800 - 0900	STATISTICAL METHODS IN SOFTWARE ENGINEERING	Nozer D. Singpurwalla George Washington University
0900 - 0915	BREAK	Common Room
0915 - 1045	SPECIAL SESSION ON LOGISTICS Main Conference Room	Chairpersons: Malcolm S. Taylor US Army Research Laboratory MAJ Kevin M. Beam US Army TRADOC, RAND
	PROBLEMS IN ARMY LOGISTICS	COL C. Terry Chase, Deputy Director US Army Strategic Logistics Agency
	ANALYSIS OF C - 141 DEPOT MAINTENANCE ACTIVITIES	James Chrissis, Professor US Air Force Institute of Technology
	TREND DETECTION OF MILITARY SPARE PARTS DEMAND DATA	Barnard H. Bissinger, Consultant US Navy Ships Parts Control Center
1045 - 1115	POSTER SESSION PSYCHOMETRIC PRINCIPLES APPLIED TO VOCATIONAL COUNSELING DATA OBTAINED FROM ARMY VETERANS Common Room	Eugene F. Dutoit Infantry Warfighting Center and John Mobley Skinner & Associates
1045 - 1115	BREAK	Common Room
1115 - 1215	GENERAL SESSION III Main Conference Room	Chairperson: David F. Cruess Uniformed Services UHS
	INFORMATION THEORY AND STATISTICS	Jorma J. Rissanen IBM Research Center, Almaden
1215 - 1345	LUNCH	

**Thursday, 29 October 1992**

**1345 - 1515 CONTRIBUTED SESSION III**

Main Conference Room

Chairperson: Todd Jones

US Army OEC

APPROXIMATE ONE-SIDED  
TOLERANCE LIMITS FOR A MIXED  
MODEL WITH A NESTED RANDOM  
EFFECT

Mark G. Vangel

US Army Materials Technology Laboratory

DETERMINATION OF CAMOUFLAGE  
EFFECTIVENESS OF SMALL AREA  
CAMOUFLAGE COVERS BY GROUND  
OBSERVERS USING THE METHOD OF  
LIMITS

R. Leon Johnson and George Anitole

US Army Belvoir Research, Development and  
Engineering Center

Christopher J. Neubert

US Army Material Command

AN AUTOMATIC NONPARAMETRIC  
ALGORITHM FOR ESTIMATING  
MODE, NUMBER, LOCATION, AND  
STRENGTH IN MULTIVARIATE DATA

Marc N. Elliot and James R. Thompson

Rice University

**1515 - 1530 BREAK**

Common Room

**1530 - 1700 CONTRIBUTED SESSION IV**

Main Conference Room

(note dual sessions)

Chairperson: Charles Holman

US Army OEC

A NEW COURSE ON THE DESIGN OF  
SIMULATION EXPERIMENTS

Russell R. Barton

Penn State University

SAMPLE-WEIGHTED AVERAGE OF  
VECTORS

Aivars Celmins

US Army Research Laboratory

THE STRATA OF RANDOM MAPPINGS

Bernard Harris

University of Wisconsin

**1530 - 1700 CLINICAL SESSION II**

Common Room

(note dual sessions)

Chairperson: Robert J. Burge

Walter Reed Army Institute of  
Research

Discussants: John Adams

RAND

Marion R. Bryson

US Army TEXCOM

W. J. Conover

Texas Tech University

Jayaram Sethuraman

Florida State University

Henry B. Tingey

University of Delaware

PROBABILITY OF RECOGNITION  
ANALYSIS IN DEGRADED  
ENVIRONMENTS

Samuel Frost

US Army Materiel Systems Analysis Activity

JUMP CHARACTERIZATION TEST

Jerry Thomas

US Army Research Laboratory

# **The Thirty Eighth Conference on the Design of Experiments in Army Research, Development and Testing**

**RAND, Santa Monica, California**

**Friday, 30 October 1992**

0730 - 0800	CONTINENTAL BREAKFAST	Common Room
0800 - 0930	CONTRIBUTED SESSION V Main Conference Room	Chairperson: Deloris Testerman US Army TEXCOM
	THE MARKOV CHAIN SIMULATION METHOD: APPLICATIONS AND APPLICABILITY IN STATISTICAL PROBLEMS	Jayaram Sethuraman Florida State University
	ASSESSMENT OF HELICOPTER COMPONENT STATISTICAL RELIABILITY COMPUTATIONS	Donald Neal and William Mathews US Army Materials Technology Laboratory
0930 - 1000	BREAK	Common Room <b>RAND Store will be open</b>
1000 - 1130	GENERAL SESSION IV Main Conference Room	Chairperson: Barry A. Bodt US Army Research Laboratory Chairman, AMSC Subcommittee on Probability and Statistics
	OPEN MEETING OF THE PROBABILITY AND STATISTICS SUBCOMMITTEE OF THE ARMY MATHEMATICS STEERING COMMITTEE	
	CHAOTIC SYSTEMS AND STATISTICS	L. Mark Berliner Ohio State University
1130	ADJOURN	MAJ Kevin M. Beam US Army TRADOC, RAND

## **PROGRAM COMMITTEE**

Gerald R. Andersen	Robert J. Burge	Jock O. Grynovicki
Carl B. Bates	Terence M. Cronin	Carl T. Russell
Kevin M. Beam	David F. Cruess	Douglas B. Tang
Barry A. Bodt	Francis E. Dressel	Malcolm S. Taylor
Melvin Brown	Eugene F. Dutoit	Henry B. Tingey



# **MODELING AND SIMULATION IN THE MILITARY: STATISTICAL ISSUES AND OPPORTUNITIES**

**DONALD P. GAVER  
PROFESSOR OF OPERATIONS RESEARCH  
NAVAL POSTGRADUATE SCHOOL  
MONTEREY, CALIFORNIA 93943**

## **1. INTRODUCTION**

The use of modeling and simulation in the DoD has increased explosively in recent years, and such growth can be anticipated to continue. The reason is the need for credible and economical tools to assist in the organization, combination, focusing and communication of knowledge, e.g., historical and theoretical scientific information, new data, and human judgment, in order to assist decision-makers and technologists with their tasks. Another rapidly growing area involves the training of human operators, such as airplane pilots or tank operators. The use of an abstract mathematical or "computer" model in place of real-life experimentation is simply mandatory in most of the situations encountered by military decision makers: it is clearly not possible to test proposed or embryonic weapon systems in a realistic variety of actual combat environments, nor to appraise their effectiveness when they are embedded in military organizations and employed to reach operational goals. The operational test and evaluation communities of the services and at the DoD level strive to subject new systems to honest testing under field conditions to the extent that resources permit, but such tests are themselves actually physical models of true combat and may not be entirely satisfactory replicas thereof. Likewise, it is infeasible to train operators and to educate commanders entirely in the field, so

synthetic environments (models) of increasing sophistication are being devised and utilized that place the "man -or person- in the loop".

The purpose of this paper is to provide an overview of some specific current modeling situations that should provide stimulation and challenges to the general analytical, but specifically statistical and operations research, communities. The plan of the paper is as follows. First we will review the definitions of "models" and "simulation" that have been proposed by the relatively new Defense Modeling and Simulation Agency (DMSO). The latter was established by the Congress in order to coordinate the many modeling efforts proposed and being pursued in the US military. Next a brief explanation will be given of two currently active areas of modeling and simulation interest: *Cost and Operational Effectiveness Analysis*, and *Future Theater Level Modeling*. Attempts will be made to indicate the needs for statistical thinking and analysis as opportunities arise. There then follows a discussion of a number of modeling areas that have been identified and enthusiastically worked on, and others that have been less popular in the past but are likely to become of emerging interest in future. Some attempt is made to point out modeling areas that have been developed and actively explored outside the military arena, the approaches and techniques of which may be worth borrowing by military modelers. I interpolate summaries of mathematical approaches to several quite specific (sub) models that may be novel and provocative to readers. For reasons of personal interest and general concern I attempt to identify sources of variability and uncertainty throughout the discussion of specific areas and models.

It can escape no-one that the models and simulations of the types reviewed here are abstract simplifications of reality, and hence are, to some possibly extensive degree, in error. There is a justifiably active concern with under-

standing and limiting that error, while preserving the advantages of flexibility and communicability that model-based, or model-assisted, analysis provides. There is an active interest, and some healthy controversy, concerning the so-called Validation, Verification and Accreditation process and its proper and defensible definition and practice. This paper concludes with a few recent references to the literature of this subject along with some comments; no doubt many readers will have their own reactions. The last word is yet to be written.

## 2. DEFINITIONS

The reader may find it useful to see definitions of relevant terms, as provided by DMSO. These are as follows; see DoD 5000.2, Aug. 1992.

- **Model:** A physical, mathematical or otherwise logical representation of a system, entity, phenomenon or process.
- **Simulation:** A method for implementing a model over time. Also, a technique for testing, analysis, or training in which real-world systems are used or where real world and conceptual systems are reproduced by a model.

The DMSO also puts models into classes, as follows

**Computer Models:** Systems and forces and their interaction are primarily represented in computer code. There may be some human interaction with the model while it is running.

**Manned Weapon System Simulations:** Individual weapon system are modeled (e.g. by a simulator) and are typically controlled by a human operator. (e.g. SIMNET).

**Instrumented Tests and Exercises:** Actual troops, weapon systems and support systems interact in as real an environment as possible, with instrumentation being used to collect and distribute status data on the force elements. (e.g. National Training Center).

It is recognized that modeling and simulation, as briefly described above, is potentially useful in various areas of interest to the military (but, of course, also in the civilian sector). Thus: in *education, training, and military operational planning and analysis*; in *research and development* for requirements definition, engineering design support and system performance assessment; in *test and evaluation* for early operational assessment and operational test design (and outcome data analysis); in *production and logistics*, i.e., for system producibility assessment, logistics requirements and distributional procedures (stocking locations and levels, issues of replacement and repair). There are many other areas in which modeling and simulation are being conducted, and in which opportunities exist for doing so more efficiently and credibly. Some of these will be identified in later sections of this paper.

### 3. COST AND OPERATIONAL EFFECTIVENESS ANALYSIS

In the acquisition process that procures new military systems for the U.S. armed forces there are several stages. In the first of these a *mission area deficiency* is identified and the appropriate service branch examines the feasibility of removing it by modification of the use of existing systems, i.e. by changes in tactics, training, or doctrine. Operational modeling and simulation, including wargaming, clearly play an important role in this examination process. If the deficiency is perceived to persist, a formal requirement for a new system is generated that specifies the critical operating capabilities of the proposed new system; this operational need document is subject to approval at Milestone 0 of the acquisition process. If approved, a concepts generation and acquisition management process begins. The latter identifies several alternative systems concepts and initiates studies of their relative cost-effectiveness; the result is presented to a decision-making body known as the Defense Acquisition Board in

the first *Cost and Operational Effectiveness Analysis (COEA)*; this is called Milestone I. Approval at this stage launches continued analysis of the competing conceptual systems and an initial prototyping. A second COEA, created at Milestone II, again assesses the impact of the prospective changes on force effectiveness and battlefield employment; comparison of the costs and effectiveness of the systems is made so as to select one system for development. This latter COEA is coordinated with a *Test and Evaluation Master Plan (TEMP)* that specifies the actual testing of the system.

Since all systems under examination are in a conceptual state during the above process a considerable amount of modeling and simulation must be relied upon to carry out the various steps in performing a COEA. We will examine and illustrate some current practice in later examples, but first mention some basic questions and issues that are universally important.

### **Modeling Issues and Questions**

Here are some of the important issues and questions that arise when a proposed new system is to be evaluated.

- Is an appropriate and satisfactorily-validated, verified and accredited computer model or man-in-the-loop simulation available to address cost and performance issues associated with the prospective new system? Is the proposed modeling and simulation system documented, transparent and well-understood enough so as to provide reasonably trustworthy results for the specific application?
- Are appropriate data bases available for use in the model?
- Have appropriate measures of system effectiveness (operational and cost) been identified?
- Is the method of cost estimation, e.g. top-down parametric and/or bottom-up engineering sufficiently accurate? Has adequate completeness of the cost estimates been achieved?

- Have *uncertainties* in system effectiveness estimates (operational and cost) been recognized, and, to the degree possible, quantified?

The answers to the above questions are operationally binary: Yes, or No; if No then further attempts at improvement must be made, best with the aid of appropriate statistical technology and viewpoint. There remain important philosophical questions concerning the manner in which the entire enterprise is conceived and carried out in practice; healthy skepticism but a sense of realism must be balanced. Such questions are being addressed by the modeling community, e.g. the Military Operations Research Society, and by others, e.g. at Rand, cf. Paul Davis (1993). There follow brief accounts of two COEA studies that should illustrate the activities required, the difficulties, and the opportunities for statisticians and operations researchers.

#### **Army: Acquisition of New Infantry Anti-Armor Weapon System-Medium (AAWS-M)**

The Army currently fields a man-carried anti-tank weapon, Dragon, that wire-guides a missile to its target. A replacement for Dragon has been proposed; it is called Javelin and operates in a fire-and forget mode, meaning that the operator need not remain vulnerable to return fire while the missile is being guided to the target, as is the case with Dragon. The Javelin operator/gunner is still exposed during the detection, launch-processing and damage assessment phases.

Since Javelin is in the conceptual stage its performance on the battlefield must be assessed by use of models. One modeling exercise is carried out using CASTFOREM (the Combined Arms Task Force Evaluation Model), which is a two-sided event-sequenced stochastic, systemic (*not* man-in-the-loop) combat model. The situation simulated is engagement between AAWS-M teams, i.e.

Dragon, or Javelin, and a threat consisting of heavy armor (tanks and armored personnel carriers). One objective of AAWS-M is to induce infantry to dismount from the APCs, thus slowing its rate of advance.

An important aspect of the CASTFOREM modeling is the (sub)model of target acquisition imbedded therein; this is based on the so-called NVEOL Algorithm for visual acquisition. The latter is, in turn, based on certain accepted theoretical principles of human vision, but in practice has apparently given evidence of bias. Some attempt is made to introduce the effects of inter-individual random variation in the simulation but the degree of empirical validity thereby achieved is unclear. A considerable amount of target information is introduced into the model, such as the target vehicle's critical dimensions, its optical and thermal contrasts associated with time of day and year, aspect, etc. Although these, and other such steps are reassuring it is still not clear how well the total physical environment, including operator effect, is represented in the portrayal of the operational difference between AAWS-M alternatives. It appears that there are opportunities for additional careful statistical investigations to be undertaken in cooperation with other scientists and test personnel in this area. It should be recognized that other complementary and supplementary studies are made using different models, such as JANUS (which is man-in-the-loop) and VIC, and that these in turn must be supplemented by cost assessments in studies to address high-level issues, such as tradeoffs between AAWS-M and heavy anti-tank weapons and the use of air-defense weaponry in an anti-armor role. The challenge to model makers, adapters, and critics, e.g. statisticians, is real and will continue.

## Navy: Vertical Take-Off and Landing (VTOL) Unmanned Airborne Vehicle (UAV)

An exciting aspect of future U.S. military forces is that they may well consist of or be supplemented by unmanned robotic elements. Such items can go where manned vehicles cannot, do not put humans at risk, and are small and difficult to detect and target. They must provide cost-effective services, e.g. information of use to force commanders.

A COEA now (1993) in progress of a prospective Navy VTOL UAV, to be deployed on small combatants in a task force, contemplates the following missions: reconnaissance, surveillance, target acquisition, deceptive ECM, (decoying of anti-ship missiles), and damage assessment. Its measures of *performance* are payload, range, endurance, speed, altitude, survivability (as affected by radiation signatures), non-combat operational attrition rate, and achievable sortie rate, among others. Measures of *system effectiveness* should include detection rate, percent of targets identified (in addition to those by, or in place of, other sensors such as manned helicopters or radars), and mission success rate in complex operational environments.

A COEA of such a future device must necessarily be conducted by simulation, using presumed actions and scenarios. An experimental simulation has been formulated within a DARPA-espoused program: Synthetic Environments for Requirements and Concepts Evaluation and Synthesis (SERCES). The notion is to link existing man-in-the-loop simulators: *Resa* located at NRAD (Once NOSC) in San Diego, where task force operational data is generated, with *Simnet* in Reston, VA., where a UAV development facility is located. Navy tactical action officers, commanders, and VTOL UAV operators interact to assess the operational impact of information. The question: given a



synthetic UAV visualization of (simulated) combat over a several-day period, did the UAV make a worthwhile contribution to combat outcome, as compared to the situation without that asset? In the network of these two simulators *Resa* provides the operational battle context and *Simnet* provides a detailed picture of what the UAV can see of the battle.

Statisticians will recognize the challenges of the experimental design and data analysis problems that rise here. A foremost problem is that of dealing with the properties of human operators and decision-makers: their inter-individual differences plus the changes and learning that occur as a result of experience with different scenarios. Also, scenario choice only initiates chains of events that can end quite differently, depending upon human operator behavior (detections, decisions) and intervening random events. Certainly the ideas of blocking and paired comparisons should be important in design and analysis, but the differences between human capacities to adapt and accommodate to new challenges may overcome apparent advantages that could lead to improved combat outcomes.

In summary, the design of trustworthy and defensible COEAs is a new challenge to statisticians who work on military problems. The challenges stem from the need to work with complex simulation tools that must be reasonably validated when linked, and that often include human operators, with their capabilities but occasional vagaries, as components.

#### **4. MODERN THEATER-LEVEL MODELS**

The theater-level models that were appropriate for planning conflict between NATO and Warsaw Pact forces are no longer suitable for anticipated conflicts of the future. Movement and maneuver, that are treated by traditional models, such as TACWAR, in a stylized piston fashion that ignore uncertainties

as to ground truth, are now described as occurring over node-arc summaries of the relevant terrain. Serious attempts are being made to model the impact of (imperfect) information concerning opponent status and intention on that maneuver; this requires an attempt to model the acquisition, dissemination, and employment of that information for force direction – in short, an attempt to model the impact of C3/I. The use of deception and dis-information will also be modeled. A strong motive will be to comprehend the special advantages of, and problems with, joint force operations.

Statisticians and operations research analysts will be challenged by the problems of creating such models; their contributions can be directly useful in modeling the various uncertainties that are realistically present, e.g. in raw sensor inputs and in the fusion thereof. They can also participate in representing the decisions that evolve from the error-prone data, and in the effects of the ultimate conflicts that either take place or are avoided, for a valuable introduction to this area, see Hillestad, Moore, and Larson (1992), and also Youngren (1992); the latter is unfortunately an unpublished manuscript.

## **5. MILITARY MODEL TYPES**

In this section we provide an inventory and description of a number of the military situations and phenomena that require models, and the types of models that are currently used, or are potentially useful, for analyzing those situations and problem areas. An attempt will be made to point out new modeling and analysis needs as a stimulus to statisticians and operations researchers, and also to research program directors; many of the needs identified are enhancements of existing models or suggestions for combining existing (sub)models into more comprehensive structures. Here are some categories of models that seem worth attention.

**Attrition.** This area is extremely broad and classical. The military connotation of the term is that of destruction or erosion of the physical force of one opponent, both human and materiel, by the force of the other. Possibly the first and most familiar models for classical force-on-force attrition are the famous Lanchester Laws, see Lanchester (1914). These are systems of ordinary first-order differential equations that resemble the rate equations of chemical kinetics and the predator-prey equations of mathematical ecology, see Beltrami (1987). Originally the equations identified just two homogeneous opposing forces, Blue and Red, but current formulations allow for different types of attrition forces on each side and account for their different attrition capabilities against each other; see Karr (1983), Anderson (1989), Taylor (1983). There is an opportunity also to model the control and coordination of such multi-type attrition forces with the assistance of C3/I assets; this opportunity has not been widely seized.

It was soon well-recognized that the original Lanchester equations were candidates for a formulational face-lift, and many modifications were proposed beginning around the end of World War II, see Morse and Kimball (1951). An initial perceived defect was the determinism of the differential equation solutions since actual combat outcomes were widely believed to appear "stochastic" or "random", i.e. not uniquely and simply related to initial force levels and the physical capabilities of one side to kill the other; this research opportunity has been addressed by formulating attrition probabilistically, e.g. as a bivariate Markov chain in continuous time with known transition function or generator. Numerical solutions to such problems can be constructed and provide insights: one is that under realistic assumptions concerning conditions of combat termination the expected values of post-combat force sizes computed from such models do not differ much from the solutions of the corresponding deterministic

equations. This can be true in an asymptotic sense; see Section 6 of this paper, but tends to ignore the effect of imperfect information and uncertainty on the decision of one side to surrender or withdraw, or to exercise other operational options. In particular it also ignores the effect of an attempt to make such a decision with the aid of a calculation of *risk*, which would require an estimate of the complete probability distribution of future casualties; such a feature will appear in the models of the future. The stochastic Lanchesterian models available also do not recognize the possibility of "double-stochasticity", meaning important and random-like variation of the attrition parameters themselves, such as rate of target acquisition, rate of fire, and kill probability. The introduction of such could economically represent terrain and general battlefield inhomogeneities. Explicit modeling of doubly stochastic effects might well provide the basis for incorporating Bayesian control into combat models. The effect of such random parameter variation could be explored by simulation, but also by mathematical methods; see Freidlin and Wentzell (1984).

Early Lanchesterian models were strictly attrition-focussed and did not explicitly account for the interaction of relative attrition and the possession of territory; however see Koopman (1963). Some attempt has been made sporadically to introduce the effect of information, via C3/I resources, on combat outcome. The subtle and intriguing area that includes information acquisition and utilization (and denial), maneuver, and ultimately the option of combat, deserves far more research attention than it appears to have received, since loss of C3/I capability can shatter modern combat capability as effectively as the loss of primary weaponry.

Still other deficiencies in the classical formulations that are currently coming under study are representations of decisions to reinforce or withdraw based on

*uncertain perceptions* of the opposition's current strength. Under some circumstances simple discrete-time (difference equation) versions of the Lanchesterian scheme, but with reinforcement that is triggered by relative force perception and is lagged in time, can generate solutions that depend non-monotonically upon initial force sizes in an unintuitive way; see Dewar, Gillogly, and Juncosa (1991). This effect can apparently be reduced by smoothing the time rate of reinforcement, but is generally disturbing because it suggests the possible dangers of assembling larger-scaled, e.g. theater-level, models from incompletely-understood submodels.

**Information: Sources.** These are other features of attrition modeling and Usage. Models that describe the integration of force elements in the style of modern theater-level combat must include accounts of information available to the respective commanders. Such information is obtained from various sensors such as satellites, manned reconnaissance flights and missions, and AUV's and GUV's. The raw, imperfect, and time-delayed (and enemy-corrupted) information as to the location, identities, speeds and directions of advance and apparent intentions of various opponent force units must be integrated ("fused") to generate a commander's perceptions that assist him to direct his own forces' actions. For the purpose of modeling force interactions it is not necessary, or desirable, to model sensor behavior in detail at the engineering "bandwidth-bit-byte" level, but rather to represent it from an operational perspective that relates measures of reconnaissance effort and coverage to probability of detection, correct classification as to unit type and asset portfolio, and the sizes of the various asset types in that portfolio. Realistic uncertainties and errors can be represented probabilistically; it seems to often be convenient to take a Bayesian viewpoint when endeavoring to combine the various information elements coherently.

There is precedent for such at the tactical level, e.g. to design tracking algorithms based on recursive up-dating ("Kalman filters") as well as the searching procedures classically described by Koopman (1980), and subsequently by many others. The present challenge is to link models of information acquisition and evaluation to models for maneuver and attrition so as to depict theater-level combat in a meaningful way.

**Logistics.** Models of demand, and re-supply, for spare parts and sub-systems, and also for consumables such as fuel and ammunition, have long been furnished by statisticians and operations researchers to military clients. The needs for such models, and the decision rules based on them, remain strong because of the current trend towards a smaller and more cost-effective armed force. Among the various universal issues is that of economically choosing the appropriate mix of repair parts, locating them geographically, and replacing them as they are consumed, taking account of the manner in which they should be used in practice—to maintain the mission availability of groups of platforms (aircraft, tanks) being supported. It is recognized that prediction of usage is difficult, so development of models that represent the ingenious adaptive processes often used by the best field logisticians is required; these processes include "cannibalization" and use of dynamic priority rules for repair and geographical distribution of spare parts. Also, ways to improve the efficiency of depot repair are of strong current interest; see Abell, Miller, Neumann, and Payne (1992). An area of current interest in the military maintenance and repair community is that of the advisability of *subsystem upgrade*, i.e. re-engineering or replacement *in the face of uncertainty*: The situation is generally as follows: if a subsystem of a major system begins to show signs of increasing unreliability and cost of maintenance, should it be upgraded, or replaced, considering that such a

change entails a substantial fixed cost and that the parent system may itself be scheduled for replacement after a definite time horizon? A related issue concerns the pros and cons of paying for stretch-out of the life of a current system instead of the procurement of a replacement for it. The fact that decisions must be made in an atmosphere of uncertainty, e.g. as to the magnitude of a newly-detected trend towards higher maintenance costs in the system-upgrade situation, should stimulate statisticians and operations researchers; see Gaver and Jacobs (1992) for an initial attack on the up-grade problem. In that work the occurrence of an adverse trend was treated as a *change-point problem*, (for a recent review of such, see Carlin, Gelfand and Smith (1992)) and uncertainties were assessed both by bootstrapping and by the invocation of a prior for the changepoint.

It is clear that a failure to maintain adequate logistics support for a combat unit may well lead to the defeat of that unit. This suggests that targetting logistics may be profitable, which in turn suggests that a realistic feature of a modern theater-level model, as discussed in Section 4, should be its dependence upon adequate supplies, and the vulnerability of those supplies to an opponent's actions.

### **Medically Oriented Models: Battlefield Casualty Management, and Environmental Toxicology**

The above-mentioned areas are important in practice but have been somewhat under-represented in the military statistics and modeling literature. The management of battlefield casualties involves *triage*, i.e. the process of making the decision as to the advantageous organizational level at which to treat an incoming combat casualty. Trade-off issues: casualties treated at the field level may be more readily returned to combat, provided they recover quickly, than if they are evacuated; on the other hand they occupy facilities and use resources,

such as hospital beds, that might be better employed for the benefit of others; for discussion and an explicit model and data analysis see T. Howard (1993). This area, which seems related to that of general emergency medical care planning and scheduling, offers many opportunities for applications of statistical and operations research techniques.

*Toxicology* is concerned with the study of the effects of "poisons" on organisms, importantly the human body and mind. The poisons of interest are typically toxic chemicals that may be encountered in the workplace or general environment in which we live by way of inhalation or skin contact, or in food and drinking water. The effects of such vary considerably depending upon the actual dosage of particular chemicals received by individuals and the individual susceptibility of the recipients; the latter may be affected by personal habits such as smoking. Statistical issues arise when exposure and dose-response relations are to be quantified. An important current issue relates to assessment of cleanup of groundwater, where the toxic condition of the latter prior to cleanup action is the result of a complex mixture of chemicals. An option for assessing the reduction of toxicity achieved by the cleanup procedure is to expose animals or fish to samples of the original water and to the water after cleanup; evidence of toxic effect, such as prevalence of liver neoplasms among those exposed to the untreated as compared to those exposed to the treated, water is used as an indicator of the effectiveness of the treatment; see Gardner

Many statistical issues arise in the planning of such studies, and the subsequent data analysis.

The above listing of topics is intended to be illustrative and stimulative; it does not pretend to be comprehensive. I next informally review some specific modeling topics that are among those of current personal interest.



## 6. STOCHASTIC MODELS FOR ATTRITION, AND FOR SURVIVAL AND RELIABILITY IN RANDOM ENVIRONMENTS

It is well-recognized that quantitative submodels used to describe military combat dynamics and their outcomes within all theater models can appropriately be formulated as non-linear, but also should be "random" or "stochastic" in one of several senses. Such challenges often cause modelers to resort to outright table-look-up to settle attrition outcomes, where the tabled values are empirically derived from historical data, or else from high-resolution submodels (ADCAL). However, for reasons of flexibility, convenience, and transparency, as well as face validity, analytical or mathematical models patterned after the classical Lanchester attrition models, cf. Taylor (1983) may be invoked to settle or predict combat outcome. The original versions of such models had many simplicities that have been, and further can be, removed in various fashions. This section reviews some such *attrition model* elaborations, emphasizing the possibilities of representing stochastic variability in the mathematical representation of attrition outcomes, thus avoiding or minimizing laborious Monte Carlo sampling.

A second area that is a candidate for continued statistical and probabilistic attention is that of *reliability and survival*, where the latter topic is interpreted broadly. In the second part of this section we describe the possible effect of *environmental variability*, first on an equipment reliability model, next with extensions to other areas as well.

### **Beyond Lanchester: Stochastics, Suppression, Reinforcements, and a Caution.**

Suppose Red and Blue forces have been in combat for time  $t$ . Then a simple stochastic or probabilistic description of their mutual attrition within  $(t, t + dt)$  is

$$dR(t) = -\beta(t)B(t)dt + \sqrt{\beta(t)B(t)}dW_R(t) \quad (6.1)$$

and

$$dB(t) = -\rho(t)R(t)dt + \sqrt{\rho(t)R(t)}dW_B(t) \quad (6.2)$$

The first, (6.1), says that the decrease in  $R$  force size is a sum: first the mean effect of attrition by  $B$  in time period of duration  $dt$  (governed by an attrition rate  $\beta(t)$ ), the second is a random component, represented as normal or Gaussian with standard deviation equal to the  $\sqrt{\text{mean}} = \sqrt{\beta(t)B(t)}$ ;  $dW_R(t)$  is Gaussian white noise with mean 0 and variance  $dt$ , the increment of a standard Wiener process. The latter approximately represents the situation in which the change in the size of  $R$  force in  $(t, t + dt)$  is viewed as locally Poisson, with mean, and hence variance,  $\beta(t)B(t)$ . Clearly such an approximation will work best when  $B(t)$  is relatively large, i.e., when the initial force  $B(0)$  is large and  $t$  is not too long, so that neither  $B(t)$  nor  $R(t)$  are close to zero; according to usual doctrine one or the other force will withdraw if losses are great enough, or its perceived force ratio becomes sufficiently unfavorable.

A discrete-time simulation of this model can easily be performed: split time into intervals of length  $\Delta$  (e.g., hours or days); then calculate force sizes  $R(\Delta)$ ,  $R(2\Delta)$ ,  $R(3\Delta)$ , starting from  $R(0)$  by evaluating the recursions

$$R(t + \Delta) = R(t) - \beta(t)B(t)\Delta + \sqrt{\beta(t)B(t)}\Delta W_R \quad (6.3)$$

where  $\Delta W_R$  is normal with mean zero and variance  $\Delta$ ; a complementary recursion holds for  $B$ , with  $\Delta W_B$  and  $\Delta W_R$  initially independent. Such a simulation is much quicker to perform than is one for a simple pure death Markov process replacement for deterministic Lanchester. Furthermore, the effect of making attrition rates  $\beta(t)$  and  $\rho(t)$  stochastic processes, e.g. so as to reflect environmental variation, is easily traced.

It is possible to justify (6.1) and (6.2) mathematically by initially specifying  $(R(t), B(t))$  as Markov with transition rates or generator

$$\begin{aligned} P\{R(t+dt) = R(t) - 1, B(t+dt) = B(t) | R(t), B(t)\} &= \beta(t)B(t)dt + o(dt) \\ P\{R(t+dt) = R(t), B(t+dt) = B(t) - 1 | R(t), B(t)\} &= \rho(t)R(t)dt + o(dt) \end{aligned} \quad (6.4)$$

and

$$P\{R(t+dt) = R(t), B(t+dt) = B(t) | R(t), B(t)\} = 1 - [\beta(t)B(t) + \rho(t)R(t)]dt + o(dt) \quad (6.5)$$

and introducing the normalized variables

$$\begin{aligned} X(t) &= [R(t) - ar(t)] / \sqrt{a} \\ Y(t) &= [B(t) - ab(t)] / \sqrt{a} \end{aligned} \quad (6.6)$$

for  $a \gg 1$ , e.g.  $a = B(0) + R(0)$ . If these are introduced into characteristic function (ch. fcn.) equations for  $(R(t), B(t))$ , i.e. Fourier transforms of forward Kolmogorov equations, and orders of  $a$  identified then there results

$$\frac{dr}{dt} = -\beta(t)b(t), \text{ and } \frac{db}{dt} = -\rho(t)r(t) \quad (6.7)$$

essentially the deterministic Lanchester square-law equations. Additionally, the limiting ( $a \rightarrow \infty$ ) ch. fcn. differential equation is that of a bivariate Ornstein-Uhlenbeck process, from which the variance-covariance structure of the random variation of  $(R(t), B(t))$  at any time  $t$  can be deduced; furthermore  $R(t)$  and  $B(t)$  are approximately jointly normally distributed. The above allows *analytical* calculation of the probability distribution of the force advantage of one side vs. the other at time  $t$  after combat begins, e.g. the advantage of  $R$  over  $B$  is  $R(t) - B(t)$  in terms of initial force sizes and attrition rates, thus providing a decision-maker with an appraisal of combat risk that augments simple mean comparisons.

Analysis of historical battles, e.g. by Hartley (1990, 1991), suggests that attrition rate is better described by a non-linear function: replace  $\beta(t)B(t)$  in (6.7)

by a suitable function of both states  $\beta[B(t), R(t), t]$ . To the extent that such a function itself has a random component, calculations may still be possible. Invocation of small random perturbation theory, e.g. Freidlin and Wentzell (1984) appears suitable.

The above analytical approach can be extended to account for other realistic operational effects. The first of these is *suppression*, wherein the effect of opponent, e.g.  $B$ , fire is to render certain forces *temporarily* incapable of active fire themselves; restoration may occur after a time. Thus the dynamics become expressed in terms of the state variables  $R_A(t)$ , the number of those active,  $R_S(t)$ , the number of those suppressed, along with  $B_A(t)$  and  $B_S(t)$ . The equations become, first for active  $R$ :

$$\begin{aligned} dR_A(t) = & -\beta_A(t)B_A(t)dt - \beta_S(t)B_A(t)dt + \mu_{RSA}(t)R_S(t)dt \\ & + \sqrt{\beta_A(t)B_A(t)}dW_{RA}(t) + \sqrt{\beta_S(t)B_A(t)}dW_{RAS}(t) \\ & + \sqrt{\mu_{RSA}(t)R_S(t)}dW_{RSA}(t) \end{aligned} \quad (6.8)$$

with a complementary setup for  $B$ . For the suppressed  $R$ ,

$$\begin{aligned} dR_S(t) = & \beta_S(t)B_A(t)dt - \mu_{RSA}R_S(t)dt + \sqrt{\beta_S(t)B_A(t)}dW_{RAS}(t) \\ & + \sqrt{\mu_{RSA}R_S(t)}dW_{RSA}(t) \end{aligned} \quad (6.9)$$

A second effect is that of *reinforcement*; a complement is *withdrawal*. First, a simple generic reinforcement model is

$$\begin{aligned} dR(t) = & -\beta(t)B(t)dt + \vartheta_R(t)H(R(t-\tau), B(t-\tau), t-\tau)dt \\ & + \text{STOCHASTIC TERM.} \end{aligned}$$

Here it is assumed that reinforcement rate at time  $t$ , portrayed by  $\vartheta_R H$ , is controlled by system state at previous, lagged, time  $t-\tau$ ; more general and realistic is a distributed lag model. Two versions of  $H$  are a step function of the form

$$H(R(t), B(t)) = \begin{cases} 1 & \text{if } B(t) / R(t) > \bar{h} \\ 0 & \text{otherwise.} \end{cases} \quad (6.10)$$

This calls for Red reinforcement at rate  $\vartheta_R(t)$  if the Blue to Red force ratio becomes sufficiently unfavorable ( $> \bar{h}$ ); otherwise none. Alternatively, smoothly increasing function may be more realistic, e.g.

$$H(R(t), B(t)) = \frac{(B(t) / R(t))^p}{(\bar{h})^p + (B(t) / R(t))^p}, \text{ for } 1 < p = o(a) \quad (6.11)$$

Use of the latter facilitates the asymptotics described earlier. Withdrawal models may be similarly constructed.

Work by Dewar, Gilloghly, and Jancosa (1991) at Rand on discrete-time deterministic Lanchester systems that allow for abrupt lag-governed reinforcement suggests that a disquieting behavior of solutions may occur: an unintended and initially unsuspected non-monotonicity of "advantage" from an increase in force size of one opponent. The cause is not completely understood. Preliminary examination, by a student at the Naval Postgraduate School, indicates that tight time-concentration of a large reinforcement is associated with the effect; if the reinforcement "time" is spread out the effect tends to disappear. Discovery of the phenomenon suggests that caution is required when simple sub-models are uncritically incorporated into larger theater-level models.

### **Survival and Reliability**

Situations that involve *survival* occur widely in both military and civilian applications. Military forces attempt to threaten or reduce the survival of members and physical assets of the opposition while guarding their own. An essential part of this activity is the detection and identification and location of opponent force elements by use of C3I assets while surviving such efforts by the enemy. Since sophisticated equipment is often involved, its survival or *reliability* is important. The survival of both military and civilian populations when

confronted with environmentally transported chemical pollution is of growing concern. There are other examples. A common mathematical thread links some of these.

Probability models have traditionally been applied to described the time to failure of equipment components and systems: if  $T_S$  represents the random lifetime of an equipment under operational use then  $\bar{F}_{T_S}(t) = 1 - F_{T_S}(t) = P\{T_S > t\}$  is commonly called the *survivor function*. Modeling a system survivor function in terms of its components should be done in terms of the system structure, cf. Barlow and Proschan (1965); the latter reflects aspects such as redundancy. Typically the modeling assumes that components fail independently in accordance with given distribution functions. Not infrequently the simple exponential is used in applied work.

Such models have been useful, and are used, yet they fail to reflect both individual and environmental components of variation that may affect failure rate parameters, and hence the distribution of failure times. Consider a series system of  $n$  components that, for the moment, are viewed as nominally indistinguishable in failure propensity. In addition each component of the system is exposed to a common but randomly fluctuating stress regime. Then the probability that the system survives for time  $t$  may be modeled as follows:

$$P\{T_S > t | \lambda_1, \lambda_2, \dots, \lambda_n, \Lambda(t)\} = \prod_{i=1}^n \exp[-(\lambda_i t + \Lambda(t))] \quad (6.12)$$

where  $\lambda_i$  ( $i=1,2,\dots,n$ ) are independent with distribution/density  $F(t)$  ( $f(t)$ ) and  $\Lambda(t)$  is the cumulative environmental hazard at time  $t$ . Then unconditionally

$$P\{T_S > t\} = (\hat{f}_\lambda(t))^n E[\exp(-n\Lambda(t))] \quad (6.13)$$

If  $\Lambda(t)$  is infinitely divisible it can be realized as compound Poisson, i.e. the varying environment occurs as a series of shocks. Consequently

$$E[\exp(-n\Lambda(t))] = \exp(-\vartheta t(1 - \hat{g}(n))) \quad (6.14)$$

with  $\vartheta$  the shock rate and  $\hat{g}$  the Laplace transform of shock magnitude. One alternative is that  $\Lambda(t)$  be a gamma process, with

$$P\{\Lambda(t) \leq x\} = \int_0^x e^{-\alpha u} \frac{(\alpha u)^{\beta t - 1}}{\Gamma(\beta t)} \alpha du \quad (6.15)$$

in which case, putting  $\beta = \vartheta$ ,

$$E[\exp(-n\Lambda(t))] = \exp(-\beta t \ln(1 + n / \alpha)) \quad (6.16)$$

Note that for the above model of environmental variation the environmental hazard component is linear in  $t$  but sublinear in  $n$ ; conventional models that do not recognize environmental variability exhibit hazard dependence linear in  $n$ .

Examine next the effect of the individual variation, expressed by  $\hat{f}_\lambda(t)$ . If  $\lambda$  comes from a gamma population,

$$f_\lambda(x) = e^{-ax} (ax)^{b-1} / \Gamma(b) \quad (6.17)$$

then the survival probability takes the form

$$P\{T_S > t\} = (1 + t/a)^{-bn} (1 + n/\alpha)^{-\beta t} \quad (6.18)$$

so here the individual hazard component is linear in  $n$  and sublinear in  $t$ , again deviating from the behavior of conventional models. The particular skew symmetry of the dependence on  $n$  and  $t$  in the formula (6.18) is a consequence of the choice of gamma variation. Note that still greater sublinearity of dependence upon  $n$  in the environmental hazard component can be obtained by replacing  $\beta t$  in (6.15) by the subordinating process  $\beta(t)$ . If the latter is again made gamma (6.16) becomes

$$E[e^{-n\Lambda(t)}] = \exp(-\beta t [\ln(1 + \ln(1 + n/\alpha)/\alpha)]) \quad (6.19)$$

The hazard thus exhibits a dependence on  $n$  that increases like  $\ln \ln n$  rather than  $\ln n$ , certainly far more slowly than that of the conventional model, which is at rate  $n$ . Of course such models are hypothetical and illustrative only, but serve to indicate the type of qualitative behavior that may broadly occur in weakest-link series systems when conventional models are plausibly modified.

Final comment: the previously-described model may be appropriate for representing the *detection-survival time of an approaching target by a system of similar (not identical) sensors, operating in an environment of varying visibility*. The model shows that survival probability does not decrease geometrically with number of sensors,  $n$ , as would be true with simple independent constant-environment models. The reason is that all sensors discussed are sensitive to the same environmental fluctuations. For improved performance a varied portfolio of sensor performance sensitivities is required.

## 7. THE VERIFICATION, VALIDATION, AND ACCREDITATION ("VV&A") OF MODELS

Since all models, no matter how complex they may be, are approximations to reality it is natural to question the adequacy of the representation provided by a particular model for the specific purposes intended. The current terminology for such general questioning activity has come to be known as VV&A, or *verification, validation, and accreditation*. As models are increasingly viewed as attractive supplements or alternatives to either unassisted expert opinion or judgement, or to excessively costly — even practically infeasible — field experiments, concern for workable and defensible definitions and procedures has intensified. Attractive as they may be, model-based analyses must often be



defended. The purpose of this section is to point out some of the current thinking and literature on the subject without attempting to be encyclopedic. For a systematic and current review of the subject the reader should consult a forthcoming Military Operations Research Society (MORS) monograph edited by Dr. Adelia Ritchie. For a currently available discussion see Davis (1992); the latter provides an overall perspective, some useful check-list and sanity-check reminders, plus good historical references.

Very informally, *verification* refers to the determination that a model realization, i.e. as a collection of interlinked numerical algorithms and submodels or modules in the form of a computer program, actually represents the the intention of the modeler. In the verification process the modeler, and often an independent party, reviews for relevance and accuracy the subject-matter information, logic, and science behind the equations and algorithms in the model; ideally, the details of this examination should be adequately documented but this ideal may not always be achieved. Simplifying assumptions made for convenience should be revealed. Then, the techniques used to solve equations and exercise algorithms should be made explicit and justified; this includes specification of any pseudo random number generation processes or Monte Carlo devices or "swindles" used to reduce sample size, as well as statistical methods used to specify point-estimates and their uncertainties, e.g. by standard errors and confidence limits. Suspicion should be aroused if simulation estimates are *not* equipped with some estimates of their uncertainty, or if the methodologies used appear inappropriate (e.g. because of unwarranted assumption of independence, or dependence on particular distributions without sensitivity check). A natural and useful verification tool or approach is to check the model's output for special cases in which the results can be calculated

independently, ideally in nearly closed form, or "on the back of an envelope". Clearly there are many tasks for statistical and operations research scientists in the verification process, not the least of which is to verify that appropriate data can be found to evaluate a model's parameters. In many cases such data may not be based entirely on physical observations but will inevitably emerge, at least in part, from expert judgement. Statistical attitudes

Again informally, *validation* aims to ensure that, given suitable initial conditions and parameter values, a model can produce predictions that agree satisfactorily with real-world outcomes. The meaning of "satisfactorily" has been expanded to "adequate for the purposes of the study of which it is part " by Miser and Quade (1988) quoted by Hodges (1991). Hodges chooses the words *bad model* to refer to a model that has not been tested and found adequate, or validated in traditional scientific fashion, i.e. by empirical comparison of predictions to relevant observational data. Presumably one that has been so tested and found adequate in some context is *good* (in present-day street parlance such, if they can be found, are called *baaad*). The dichotomous labelling is convenient, but is itself an abstraction for something more diffuse and vague that must be expanded upon and made specific if the potential user is to consider dependence upon model implications. It seems conceivable that a model labelled *bad* given certain exposures to reality, or lack thereof, could become *good* under others, and the reverse. Thus it may be desirable to try for an economical description of model credibility status that provides more information to a prospective user or client than a simple stamp of approval/disapproval. It is well-recognized that many useful military models, particularly those of engagement or combat using conjectural future technology, can never be validated in a totally empirical-scientific way. But some such have more face validity than others or were

constructed by credible craftspersons and thus may have a kind of genetic credential. In short I contend that all models that are not empirically validated are not equally dubious; many are quite useful to-- and used by-- responsible decision-makers. Since they will wish to appraise and evaluate any models, users and analysts are deserving of, and grateful for, careful and complete documentation at a technical level.

Note that frustration with model validation is not confined to military combat modelers but, for example, plagues those concerned with acquiring and positioning appropriate levels of logistics in a theater, with planning for the medical treatment of combat casualties, and with assessing the health effects of environmental pollution. Perhaps in time a generally accepted approach to the validation questions will appear. It seems possible that the statistical literature on inference from *observational data*, as expounded by Cochran and D. Rubin, is relevant.

The decision that a particular model is suitable for a given purpose, i.e. for supplementing field tests in operational test and evaluations or particularly in COEA, is called *accreditation*. In order for a model to be credibly accredited, the model itself must be well-understood and verified and validated ( to the degree possible), and it must be capable of answering questions relevant to the particular application. It should be helpful to know that the model has been previously used for a given purpose similar to the present one (e.g. missile survivability or reliability testing), and that the model's predictions were similar to later real-world experience; such information may not be readily available, either because it does not exist, or perhaps because it is unknown to the agency responsible for carrying out the present application. A handy compendium of

models that have been used with success in certain classes of applications could  
be very useful.

## BIBLIOGRAPHY

- Abell, J. B., Miller, L. W., Neumann, C. E., and Payne, J. E. (1992). *DRIVE (Distribution and Repair in Variable Environments). Enhancing the Responsiveness of Depot Repair*. RAND (R-3888-AF). Santa Monica, CA.
- Anderson, L. B. (1989). *Heterogeneous Point Fire and Area Fire Attrition Processes that Explicitly Consider Various Types of Munitions and Levels of Coordination*. IDA Paper P-2249, Institute of Defense Analyses. Alexandria, VA.
- Anderson, L. B., and Miercort, F. A. (1989). *COMBAT: A Computer Program to Investigate Aimed Fire Attrition Equations, Allocations of Fire, and the Calculation of Weapons Scores*. IDA Paper P-2248, Institute of Defense Analyses. Alexandria, VA.
- Barlow, R. E., and Proschan, F. (1965). *Mathematical Theory of Reliability*. SIAM series in applied mathematics. New York, NY.
- Beltrami, E. (1987). *Mathematics for Dynamic Modeling*. Academic Press, Inc. San Diego, CA.
- Davis, P. K. (1992). *Generalizing Concepts and Methods of Verification, Validation, and Accreditation (VV&A) for Military Simulations*. RAND (R-4249-ACQ). Santa Monica, CA.
- Dewar, J. A., Gillogly, J. J., and Jancosa, M. L. (19 ). *Non-Monotonicity, Chaos, and Combat Models*. RAND (R-3995-RC). Santa Monica, CA.
- Freidlin, M. I., Wentzell, A. D. (1984). *Random Perturbations of Dynamical Systems*. Springer-Verlag. New York, NY.
- Gardner, H. S., van der Schalie, W. H., Wolfe, M. J., and Finch, R. A. (1990). New methods for on-site biological monitoring of effluent water quality. *In-Situ Evaluations of Environmental Pollutants*. S. S. Sandhu, et al., editors. Plenum Press. New York, NY.
- Gaver, D. P. and Jacobs, P. A. (1992). *Statistical Approaches to Detection and Quantification of a Trend With Return-On-Investment Application*. Naval Postgraduate School Technical Report (NPSOR-93-007). Naval Postgraduate School, Monterey, CA.
- Hartley, D. S. III (1990). *Historical Validation of an Attrition Model*. Martin Marietta Applied Technology, Report K/DSRD-115. Oak Ridge, TN.
- Hartley, D. S. III (1991). *Predicting Combat Effects*. Martin Marietta Applied Technology, Report K/DSRD-412. Oak Ridge, TN.
- Hodges, J. S. (1991). Six (or so) things you can do with a bad model. *Operations Research*, Vol. 39, pp. 355-365.

- Hodges, J. S., and Dewar, J. A. (1992). *Is It You or Your Model Talking? A Framework for Model Validation*. RAND (R-4114-AF/A/OSD). Santa Monica, CA.
- Howard, T. L. (1993). *An Analytical Model for the Treatment and Evacuation of Casualties in a Low-Intensity Conflict*. Naval Postgraduate School Master-of-Science in Operations Research thesis. Naval Postgraduate School, Monterey, CA.
- Karr, A. F. (1983). *Lanchester Attrition Processes and Theater-Level Combat Models*. Chapter in *Mathematics of Conflict*, North-Holland, Amsterdam, The Netherlands.
- Koopman, B. O. (1963). Analytical treatment of a war game. *Actes de la 3<sup>d</sup> Conference Internationale de Recherche Operationnelle*, Oslo, Norway. English Universities Press Ltd. London.
- Koopman, B. O. (1980). *Search and Screening: General Principles with Historical Applications*. Pergamon Press. Elmsford, NY.
- Lanchester, F. W. (1914). Aircraft in Warfare: The Dawn of the Fourth Arm—No. V, The Principle of Concentration. *Engineering* 98, p. 422–423 (Reprinted on pp. 2138–2148 of *The World of Mathematics*, Vol. IV (1956). Simon and Schuster, New York, NY.
- Miser, H. J., and Quade, E. S. (1988). Validation. In *Handbook of Systems Analysis: Craft Issues and Procedural Choices*. North-Holland. New York. pp. 527–565.
- Morse, P. McC. and Kimball, G. E. (1951). *Methods of Operations Research*. Technology Press, MIT, and John Wiley & Sons, NY.
- Taylor, J. G. (1983). *Lanchester Models of Warfare*. Military Operation Research Society of America. Arlington, VA.
- Youngren, M. A. (1992). Future theater-level model. Draft report, J-8/CFAD. The Joint Staff. The Pentagon. Washington, DC.

# **A Comparative Study of Boresight Devices for Tank Cannon**

David W. Webb  
MAJ Bruce J. Held

US Army Research Laboratory  
Weapons Technology Directorate  
Aberdeen Proving Ground, MD 21005

## **Abstract**

The use of a boresight device to align the muzzle of a tank's cannon with its fire control system is one of the fundamentals of tank gun accuracy. Various systems and devices have been used to accomplish this task, ranging from crossed strings to mark the bore centerline to precision optical devices that can cost tens of thousands of dollars. There are several optical devices currently in use with the U.S. Army's testing facilities and with line units of the army. This paper focuses on a test that compared five different types of boresight devices — from its design, to implications for improving tank gun accuracy learned during the ensuing analysis.

## **I. Introduction**

In the ongoing effort to improve the effectiveness of U.S. Army weapon systems, much time has been spent in the study of tank gun accuracy. The overall accuracy and precision of tank weaponry is influenced by many factors, one of which is the alignment of the cannon muzzle with the fire control system (FCS). The procedure by which this is accomplished is known as boresighting. The optical instrument used in the procedure is referred to as a boresight device, or simply a boresight.

To boresight the tank, the device is inserted into the muzzle of the cannon. The cannon is moved until the reticle seen through the boresight's eyepiece is centered at a target whose range is known. Then the floating reticle of the gunner's primary sight (GPS) is moved by toggle switches until it is centered onto the target also. The FCS computer then determines the azimuth and elevation angles between the centerlines of the cannon and the GPS, thus properly boresighting the tank.

Various systems and devices have been used to boresight a tank, ranging from crude crossed strings to the high-quality optical devices of today. No matter which method or device is used, ultimate accuracy depends on repeatability from occasion to occasion, even when different gunners boresight the tank. Furthermore, the current Army concept of a fleet zero, instead of an individual zero, implies that calibration with a boresight must be consistent among tanks and among devices.

An important property of any boresight is that its optics be parallel with the muzzle centerline axis. When this is true, the boresight is said to be perfectly collimated. Usually a device will be not perfectly collimated, but may be within certain specifications established by the manufacturer. When a device is "out of collimation" it must be adjusted before it may be used. A boresight that the field soldier is permitted to adjust is said to be collimatable. Otherwise, the

adjustments may only be made by qualified personnel at a maintenance facility.

To compensate for any imperfections in a boresight that may still be within collimation specifications, some manufacturers recommend a two-stage boresighting procedure. That is, after boresighting is completed, the device is rotated 180 degrees in the muzzle and a second boresighting is performed. The results of each boresighting event are then averaged to compensate for the imperfect collimation.

This presentation describes a test designed to compare the boresights of five different producers, hereafter denoted as Boresights A through E. The boresights differ in several ways such as reticle pattern and thickness, magnification, and muzzle support. Two devices from each producer were tested. The ten individual devices will be denoted by A1, A2, B1, B2, . . . , E1, and E2. In addition, comparisons were made between different devices of the same type, different gunners, and different tanks.

## II. Test Plan and Analysis -- Part I

The test was conducted in two parts. The primary objective of Part I was to obtain estimates of the total boresighting error. This error is actually a combination of several component errors including

1. the variation of boresight placement in tube,
2. changes in the geometry of the tube between readings (e.g., thermal bending),
3. the inability to make very precise movements of the cannon,
4. drift in the muzzle pointing angle between the time that the gunner completes laying the gun until angular measurements are read and recorded,
5. the variation of eye placement on the eyepiece, and
6. the inherent inability to read the same exact point on the target every time.

The test matrix for this part of the test is shown in Table 1. Six test personnel were employed as the gunners. The test was completed in three days, using two gunners each day. Each gunner boresighted each of the ten devices three times. The sixty boresighting events of each day were conducted in a completely randomized order. To broaden the range of the test, three tanks were used, one on each day, thus confounding tank effects and day effects.

The standard deviation of the three readings for each gunner/device combination was used as the measure of dispersion, or repeatability. By pooling, an overall dispersion for each boresight device was obtained as shown in Figure 1. (Note: units are withheld from all figures due to classification restrictions.)

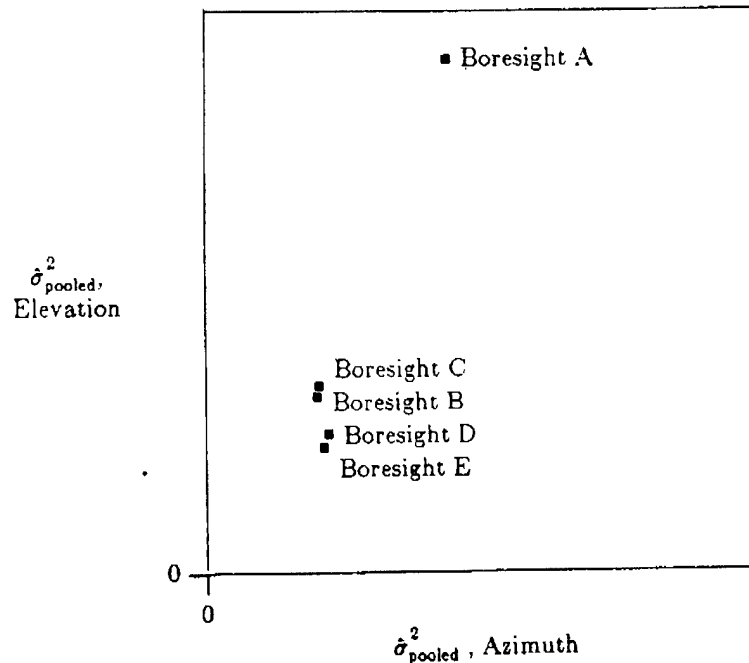
In both azimuth and elevation, the variation of Boresights B, C, D, and E is statistically the same. This variability is just slightly larger in elevation than in azimuth. Boresight A, however, has a dispersion estimate that is twice that of the other devices; and in elevation its estimate is three times that of the other devices.



Table 1. Test Matrix for Part I.

Tank/Day	Gunner	Boresight A		Boresight B		Boresight C		Boresight D		Boresight E	
		Device A1	Device A2	Device B1	Device B2	Device C1	Device C2	Device D1	Device D2	Device E1	Device E2
1	a	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)
	b	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)
2	c	same as above									
	d										
3	e	same as above									
	f										

Figure 1. Overall Dispersion of Muzzle Pointing Angle for Each Boresight Device, Part I.



A second objective of Part I was to evaluate the effect of several factors on the boresighting process. These factors were Boresight Type, Device-within-Type, Gunner, and Tank.

Simultaneous hypothesis tests for these variables and their interactions were performed through a mixed-model analysis-of-variance (ANOVA). The design treated Tank, Gunner, and Device-within-Type as random effects and Boresight Type as a fixed effect, with Device-within-Type nested under Boresight Type and Gunner nested under the blocking variable Tank.

Separate analyses were performed on the azimuth and elevation muzzle pointing angles, respectively referred to as X and Y. The expected mean squares and F (or pseudo-F) ratios appear in Table 2.

**Table 2.** Expected Mean Squares and F Ratios.

Source	df	EMS	F ratio
T [ Tank ]	2	$\sigma_e^2 + 3\sigma_{GD}^2 + 6\sigma_{TD}^2 + 30\sigma_G^2 + 60\sigma_T^2$	$\frac{MS_T + MS_{GD}}{MS_G + MS_{TD}}$
G(T) [ Gunner ]	3	$\sigma_e^2 + 3\sigma_{GD}^2 + 30\sigma_G^2$	$\frac{MS_G}{MS_{GD}}$
B [ Boresight ]	4	$\sigma_e^2 + 3\sigma_{GD}^2 + 6\sigma_{GB}^2 + 6\sigma_{TD}^2 + 12\sigma_{TB}^2 + 18\sigma_D^2 + 36\phi_B$	$\frac{MS_B + MS_{TD}}{MS_D + MS_{TB}}$
D(T) [ Device ]	5	$\sigma_e^2 + 3\sigma_{GD}^2 + 6\sigma_{TD}^2 + 18\sigma_D^2$	$\frac{MS_D}{MS_{TD}}$
T x B	8	$\sigma_e^2 + 3\sigma_{GD}^2 + 6\sigma_{GB}^2 + 6\sigma_{TD}^2 + 12\sigma_{TB}^2$	$\frac{MS_{TB} + MS_{GD}}{MS_{TD} + MS_{GB}}$
T x D(B)	10	$\sigma_e^2 + 3\sigma_{GD}^2 + 6\sigma_{TD}^2$	$\frac{MS_{TD}}{MS_{GD}}$
G(T) x B	12	$\sigma_e^2 + 3\sigma_{GD}^2 + 6\sigma_{GB}^2$	$\frac{MS_{GB}}{MS_{GD}}$
G(T) x D(B)	12	$\sigma_e^2 + 6\sigma_{GB}^2$	$\frac{MS_{GD}}{MSE}$
Error	120	$\sigma_e^2$	

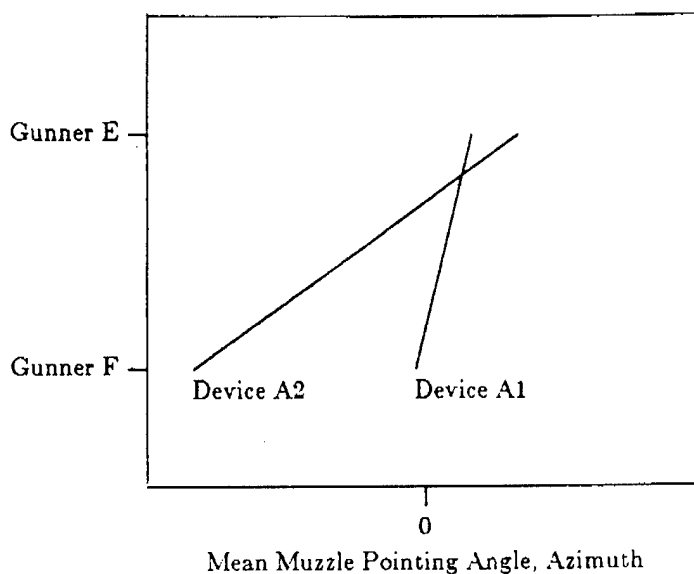
The ANOVA indicated the following significant factors at the 5% level:

- In azimuth,
  - Tank-3 Gunners \* Boresight-A Devices
  - Tank-3 Gunners \* Boresights
  - Tanks \* Boresight-A Devices
- In elevation,
  - Tank-3 Gunners \* Boresights
  - Tanks \* Boresight-A Devices
  - Tanks \* Boresight-B Devices
  - Boresights

Beginning with the azimuth, these results will be discussed. Figure 2 shows the mean azimuth readings for Tank-3 Gunners using the two Boresight-A Devices. Gunner E readings on device A1 were slightly to the left of readings on device A2. In the presence of no interaction, one would expect Gunner F to read about this same amount to the left using device A1. However the

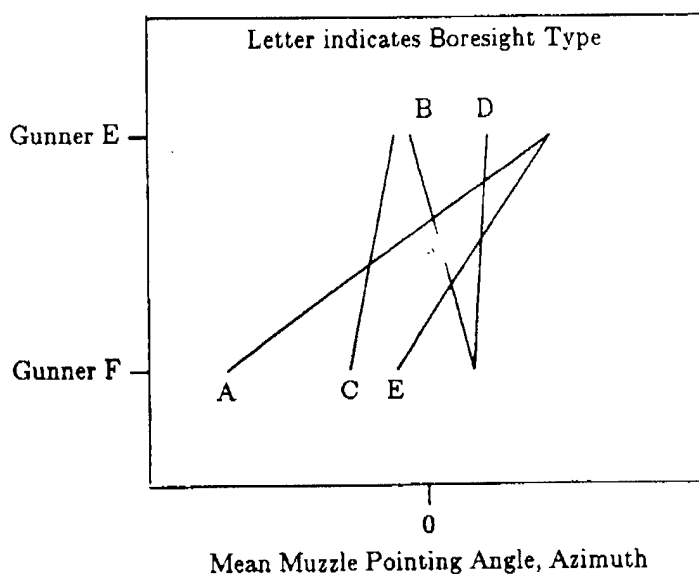
actual difference was noticeably larger and to the *right*. Therefore, the difference between the two boresight devices was dependent upon which gunner was taking the reading. This indicates that the two factors interact. Pictorially, the interaction is indicated by the non-parallelness of the two lines.

**Figure 2.** Interaction between Tank-3 Gunners and Boresight-A Devices, Azimuth.



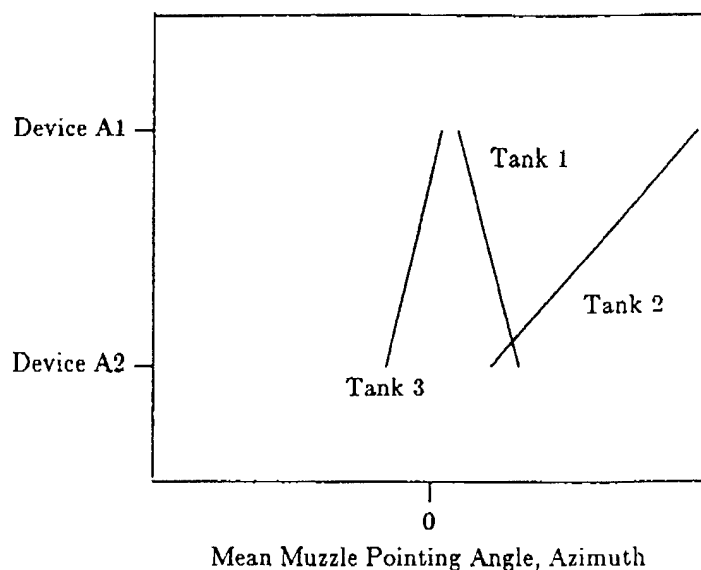
The interaction between Tank-3 Gunners and Boresights (see Figure 3) is due to the fact that the mean in azimuth readings for Gunners E and F using Boresight A was at least as twice as large as the same difference using any other Boresight Type.

**Figure 3.** Interaction between Tank-3 Gunners and Boresights, Azimuth.



A very strong interaction was also noted between Tanks and Boresight A devices. This may be seen by the the noticeable lack of parallelism of the three lines in Figure 4.

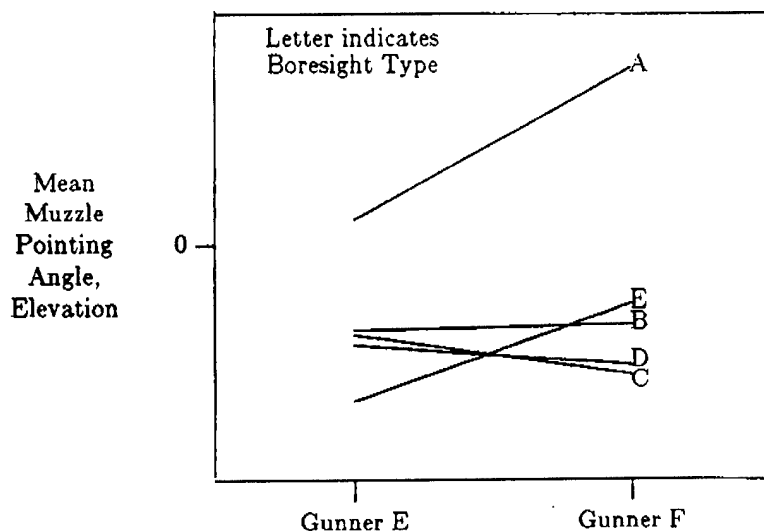
**Figure 4.** Interaction between Tanks and Boresight-A Devices, Azimuth.



Each of these three interactions found in azimuth may be explained by a faulty collimation of one of the Boresight-A devices on the third day and/or the relative instability of Boresight A.

In elevation, the ANOVA pointed out several significant effects listed previously. The first of these is a very strong interaction between Tank-3 Gunners and Boresights. This is seen in Figure 5 which depicts the elevation readings for Gunners E and F with each of the five boresight types. The difference in each gunner's mean reading is boresight dependent, therefore the significant interaction. The main contributors to this interaction are Boresights A and E. This indication of instability in Boresight E is somewhat surprising, and is possibly the result of a faulty collimation.

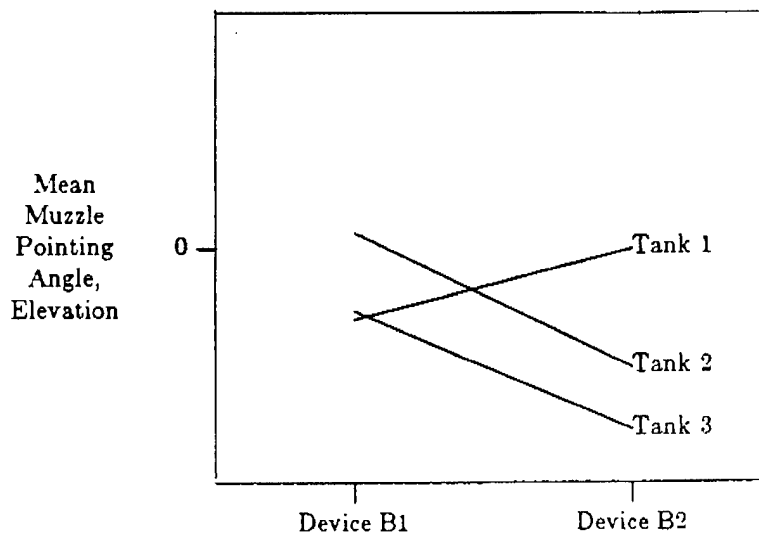
**Figure 5.** Interaction between Tank-3 Gunners and Boresights, Elevation.



Significant interactions between Tanks and Devices occurred for both Boresights A and B, as noted in Figures 6 and 7. The reason for the interaction of Figure 6 may lie in the collimation

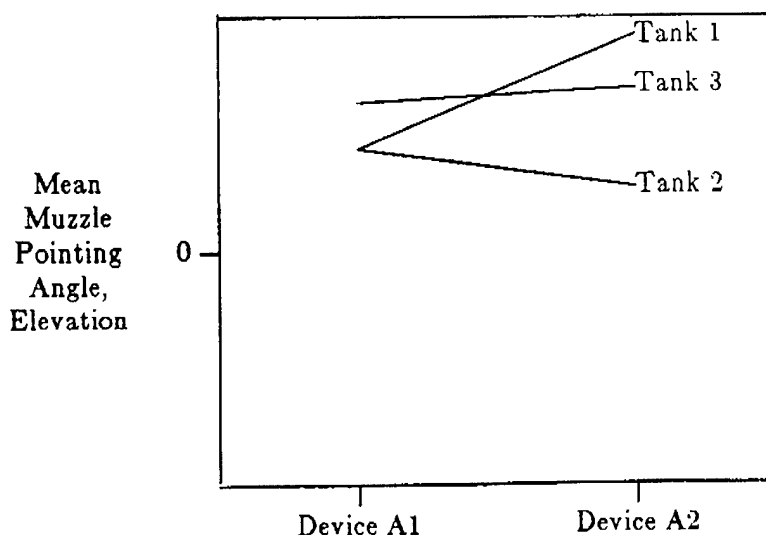
procedure. Boresight B is not required to be collimated on the same tank on which it may be later used (as are Boresights A and E). Furthermore, it does not require two readings taken from opposite sides of the tube to average any parallax error (as do Boresights C and D). Since different gun tubes wear differently, boresight devices will also sit differently in the barrel. Collimation is supposed to correct for parallax errors in tube and boresight device pairings. So if collimation is not performed, or if multiple readings are not taken, parallax errors may show up as an interaction between boresight and tank.

**Figure 6.** Interaction between Tanks and Boresight-B Devices, Elevation.



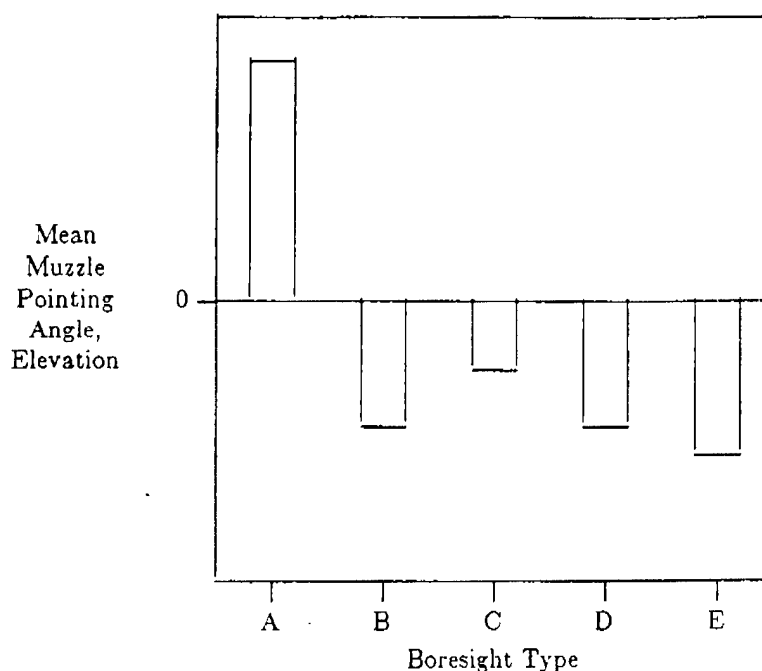
The Tank and Boresight A interaction of Figure 7 is explained again by faulty collimation and/or instability of this boresight's optics.

**Figure 7.** Interaction between Tanks and Boresight-A Devices, Elevation.



Finally, a significant boresight effect was noted. Even in light of all the interactions noted previously, Boresight A tends to read significantly higher than all of the other boresight types. Figure 8 shows the average muzzle pointing angle read from each type of boresight.

Figure 8. Mean Muzzle Pointing Angle for Each Boresight, Elevation.



### III. Test Plan and Analysis -- Part II

A second part of the overall test was designed to determine the error that is associated with the reading of the boresight. This was achieved by keeping the cannon and boresight stationary until a complete set of readings could be taken. The readings were taken in a different manner than they were in Part I. An additional test assistant stood downrange in front of a panel holding a small cross. The gunner directed this "crossbearer" via radio to move the cross until a designated corner of the cross was in line with the reticle crosshair (or dot). The crossbearer then lightly marked on the panel the position of that corner of the cross. This procedure was repeated until all five markings were made. Knowing the distance to the panel from the muzzle, the angular measurement of each reading was determined, and the dispersion of the five markings obtained.

Part II was conducted in one day using a single tank. It required four test personnel and the same ten boresighting devices that were used in Part I. The design matrix for this part of the test is shown in Table 3. The selection of a gunner/device combination was conducted in a completely randomized fashion, however all readings for that combination were taken consecutively as noted above.

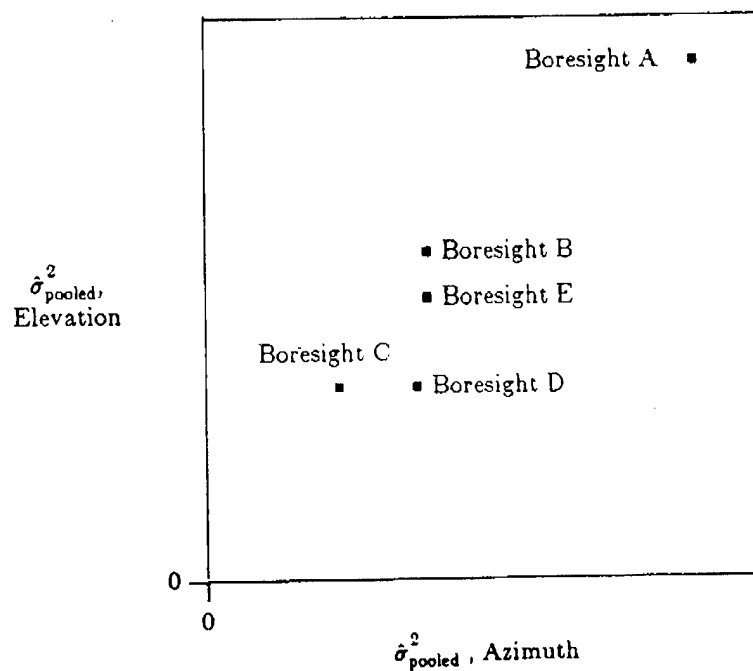
Figure 9 displays the pooled dispersions in azimuth and elevation of each boresight type. This error associated with the reading of the boresight device makes up a very small part of the total boresight error that was measured in Part I of the test. In both azimuth and elevation, Boresight A had the highest dispersion of all types tested. The larger error associated with this boresight is primarily the result of thicker reticle lines, which made it difficult to find the center of the marker cross. This effect was particularly noticeable when heat shimmer partially

obscured the target cross.

**Table 3.** Test Matrix for Part II.

Gunner	Boresight A		Boresight B		Boresight C		Boresight D		Boresight E	
	Device A1	Device A2	Device B1	Device B2	Device C1	Device C2	Device D1	Device D2	Device E1	Device E2
1	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)
	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)
	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)
	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)
	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)	(x, y)
2	same as above									
3	same as above									
4	same as above									

**Figure 9.** Overall Dispersion of Muzzle Pointing Angle for Each Boresight Device, Part II.



#### IV. Conclusions and Recommendations

- With the exception of Boresight A, all the other boresights were able to measure the same centerline of the muzzle of the cannon with good accuracy. Boresight A on the other hand, measured significantly different bore centerlines, particularly in elevation.
- Boresight A interacted with Tank 3 gunners in both directions. This is further evidence of the instability of this boresight type.
- For Boresights B and E there was some evidence of instability in elevation as these types of boresights interacted with Tanks and Tank-3 gunners, respectively. For these two boresight types, this is most likely the result of imperfect collimation. In the case of Boresight B, collimation was not authorized at the user level. Boresight E collimation was performed each morning, but improper or poorly performed collimation could lead to the noted interaction with Tank-3 gunners.
- Taking two readings, with the boresight turned 180 degrees between readings, adequately corrects for collimation errors. This should eliminate the interactions noted above with Boresights B and E.
- Boresights B, C, D and E did not differ in terms of repeatability. Boresight A as a whole had larger dispersions of its readings.
- Boresight A should not be used until it is re-evaluated. It is possible that the two used in this test were poorly maintained or were unrepresentative of the design. Until repeatability problems are resolved and it is clear that Boresight A measures the same bore centerline as the boresights in field use, its use may contribute to significant accuracy errors.



# THE RANK TRANSFORMATION IN BALANCED INCOMPLETE BLOCK DESIGNS

W. J. Conover  
College of Business Administration  
Texas Tech University  
Lubbock, Texas 79409

**ABSTRACT.** The rank transformation procedure, where the data are replaced by ranks in an overall ranking of the data and then the usual parametric procedure is computed on the ranks, is reported in this paper to be a valid procedure in balanced incomplete block designs. Computer simulation is used to see how well the F distribution approximates the exact null distribution for this procedure and four competitor procedures under five different distributional assumptions. Also computer simulation is used to compare the power of this procedure with five competitor procedures under three different distributional assumptions. The rank transformation procedure is shown to be both robust and powerful as compared with these other procedures.

**1. INTRODUCTION.** An example of a situation where a balanced incomplete block (BIB) design is appropriate is as follows. Various scents are compared, to see which are the most attractive to coyotes, in a predator-control study. Seven scents are evaluated, but a maximum of three scents can be compared at the same time, so a BIB design is appropriate. A  $7 \times 7(3,3,1)$  design is selected. The notation  $7 \times 7(3,3,1)$  refers to  $txb(k,r,\lambda)$ , where

- t = the number of treatments (seven scents in this case)
- b = the number of blocks (repetitions, where each rep compares three scents)
- k = the number of treatments compared in each rep
- r = the number of times each scent is tested (Note  $kb = rt$  in a BIB design.)
- $\lambda$  = the number of blocks where treatment i is compared with treatment j which is the same for all pairs of treatments in a BIB design. (Note  $\lambda = r(k-1)/(t-1)$  in a BIB design.)

The total time (in seconds) the coyote spends at each scent is the measure of attractiveness, and is the dependent variable. The experiment is real, but the numerical results are hypothetical, on the next page.

The classical parametric test (see e.g., Cochran and Cox, 1957), using the test proposed by Yates (1936) results in an F statistic of 1.87. This assumes normal populations, and additive block effects. Under the null hypothesis of no treatment effects this F statistic is compared with tables of the F distribution with 6 and 8 degrees of freedom. The p-value associated with  $F = 1.87$  is 0.22, which indicates there is no significant difference in scents.

		<u>SCENT</u>					
	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>
<u>COYOTE 1</u>	14	23		12			
<u>2</u>		17	2		3		
<u>3</u>			6	1		16	
<u>4</u>				0	10		42
<u>5</u>	15				4	7	
<u>6</u>		67				5	18
<u>7</u>	31		0				22

A closer look at the data reveals a few very large observations, such as possibly the observations 31 and 42, and certainly the observation 67. This in combination with the many observations less than 10 suggests that the normality assumption may not be valid. Therefore a nonparametric test is called upon.

The usual nonparametric test for BIB designs is the Durbin test, that analyzes the ranks of the observations within each block (coyote). The ranks of the observations are given as follows.

		<u>SCENT</u>					
	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>
<u>COYOTE 1</u>	2	3		1			
<u>2</u>		3	1		2		
<u>3</u>			2	1		3	
<u>4</u>				1	2		3
<u>5</u>	3				1	2	
<u>6</u>		3				1	2
<u>7</u>	3		1				2

The Durbin test statistic (see Conover, 1980) is  $D = 12$ , which is the maximum possible value of the test statistic for this BIB design because of the perfectly consistent ordering of the scents. The Durbin test statistic is asymptotically chi-squared distributed with  $t - 1 = 6$  degrees of freedom. The p-value, obtained from the chi-squared distribution, is 0.065, which is smaller than before but still not significant at the  $\alpha = 0.05$  level.

Because the rankings are perfectly consistent in the above BIB design, it is simple to compute the exact p-value of the Durbin Test statistic. The exact p-value associated with the most extreme value of D is given by

$$\frac{t!}{(k!)^b} = \frac{7!}{(3!)^7} = 0.018$$

This illustrates the fact that the Durbin test actually shows significant differences to exist among the various scents, but that the chi-squared approximation is not very good in the tail of this distribution.

An alternative procedure to the classical F test and the Durbin test is to use a rank transformation procedure (Conover and Iman, 1981). That is, all of the observations are ranked from smallest to largest, and the classical F test is performed on these overall ranks. The ranks of the observations, over all of the blocks simultaneously, are given as follows.

		<u>SCENT</u>						
		<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>
<u>COYOTE</u>	<u>1</u>	12	18		11			
	<u>2</u>		15	4		5		
	<u>3</u>			8	3		14	
	<u>4</u>				1.5	10		20
	<u>5</u>	13				6	9	
	<u>6</u>		21				7	16
	<u>7</u>	19		1.5				17

The F statistic computed on these ranks is  $F_r = 5.19$ , which is compared with the F distribution with 6 and 8 degrees of freedom, as in the classical F test, to get a p-value of 0.02. Thus the rank transformation test and the Durbin test give similar results for this set of data, while the classical F test is plagued with outliers which hamper the power of that normal-theory-based test.

This rank transformation test is not nonparametric. Is it even a valid test? What is known about the behavior and characteristics of this statistical procedure, both under the null hypothesis and under the alternative hypothesis? Specifically, the following questions need to be addressed in order to make an informed decision as to whether or not to use this rank transformation procedure in balanced incomplete block designs.

Question 1: Is the asymptotic distribution of  $F_r$  really the F distribution with  $t-1$  and  $bk-b-t+1$  degrees of freedom under the null hypothesis?

Question 2: How good is the F distribution as an approximation to the exact distribution?

Question 3: How does the power of this rank transformation test compare with its most obvious competitors?

Question 4: Can these results be extended to general scores other than ranks?

Question 5: Can the theoretical asymptotic relative efficiency (ARE) be obtained for this rank transformation test?

Question 6: Can some of these results be extended to the general linear model?

The answers to these questions have been obtained by the author as a result of research supported by the Army Research Office, and will be shown in detail in another paper. A summary of the results is given below.

2. THE NULL DISTRIBUTION OF THE TEST STATISTIC. Although it is not shown in this paper, the null distribution of the test statistic  $F_r$  is asymptotically the same F distribution used in the classical analysis, namely the F distribution with  $t-1$  degrees of freedom in the numerator and  $bk-b-t+1$  degrees of freedom in the denominator, under some easily met conditions. Specifically these conditions, in addition to the null hypothesis of no treatment effects being true, are as follows.

1. A "uniform mixing condition" holds. That is, the block effects are randomly mixed with the BIB design treatment pairings, so that some treatments don't consistently appear in blocks with higher (or lower) mean effects, which would introduce an artificial apparent treatment effect.

2. The number of blocks,  $b$ , tends to infinity.

3. The  $X_{ij}$  are independent random variables, distributed according to the block distribution function  $F_i(x)$  which must be nondegenerate in all but a finite number of blocks, and the average block distribution function

$$H_b(x) = \frac{1}{b} \sum_i F_i(x)$$

converges uniformly to some function  $H(x)$  for all  $x$  as  $b$  approaches infinity.

The next question that needs to be addressed is how well the F distribution serves as an approximation to the true distribution

of  $F_r$  when the sample sizes are small. To answer this question, a simulation study was conducted to see what percentage of the time the null hypothesis was rejected when  $F_r$  was used as a test statistic and compared with the approximate quantiles from the  $F$  distribution with  $t-1$  and  $bk-b-t+1$  degrees of freedom. Thirteen BIB designs were studied, ranging from ones with 12 observations to one with 99 observations. The number of repetitions was 2500 in each case. The block effects were additive, and the error terms were distributed according to five distributions. First the normal distribution was studied, then the lognormal, the laplace, the uniform, and finally the cauchy.

At the same time, the percentage of rejections under the null hypothesis was computed for four other tests. These include the classical  $F$  test, the aligned ranks test ART where the observations are "aligned" by subtracting the block means before an overall ranking and the  $F$  statistic is computed on the resulting ranks, the Durbin test D1 using the usual chi-squared distribution as an approximation, and a modification D2 of the Durbin test where the classical  $F$  test is conducted on the ranks within blocks.

First the results of the normal distribution simulation are presented below in Table 1. It is easily seen that the  $F$  approximation in the rank transformation test is quite good even for small sample sizes. When the total number of observations  $n = tr = bk$  is quite small the empirical estimate of the true  $\alpha$  is never larger than .058 nor less than .044, and usually quite close to the target value of .050. Of course the  $F$  test is exact, and the column of empirical Type I error rates given under  $F$  merely reflects the type of sampling variability one can expect in a simulation study of this size. The variability under the  $F$  test has the same magnitude as the variability under the RT test, suggesting that most of the variation observed under RT is due to sampling variability caused by the simulation.

Table 1. Percentage of times the null hypothesis was rejected in 2500 simulations of a BIB design, with normal random variables.

<u>BIB Design</u> txb(k, r, $\lambda$ )	<u>Statistical Test</u>				
	<u>RT</u>	<u>F</u>	<u>ART</u>	<u>D1</u>	<u>D2</u>
1. 4x6(2, 3, 1)	.052	.041	.062	.000	.000
2. 4x4(3, 3, 2)	.058	.049	.062	.000	.072
3. 5x10(2, 4, 1)	.057	.050	.062	.000	.115
4. 5x5(4, 4, 3)	.050	.053	.059	.025	.064
5. 5x10(3, 6, 3)	.056	.054	.057	.038	.060
6. 6x10(3, 5, 2)	.054	.047	.050	.028	.057
7. 6x6(5, 5, 4)	.054	.049	.056	.044	.058
8. 7x7(3, 3, 1)	.050	.045	.053	.000	.088
9. 7x7(4, 4, 2)	.049	.051	.050	.022	.048
10. 7x7(6, 6, 5)	.044	.039	.041	.032	.044
11. 8x8(7, 7, 6)	.050	.048	.049	.038	.053
12. 9x9(8, 8, 7)	.049	.054	.050	.050	.046
13. 10x10(9, 9, 8)	.050	.049	.050	.047	.045

The aligned ranks test, on the other hand, shows slightly more variation than either the RT or the F tests, with empirical  $\alpha$  levels ranging from .041 to .062.

The Durbin test with the chi-squared approximation appears to be very conservative, with no rejections reported in 4 of the 13 designs. In design 8, the  $7 \times 7(3,3,1)$  design, it is mathematically impossible to obtain a result in the rejection region, as was pointed out in the introduction, where this same design was used as an illustration. Although we did not check this out, it appears that the same may be true of designs 1, 2, and 3.

The Durbin test with the F approximation shows erratic behavior, with empirical Type I error rates as low as .000 and as high as .115.

Although the F test always performs well for normally distributed random variables because it is derived for that case, it relies on its property of robustness for nonnormal distributions. When the underlying distribution has short to moderate tails, such as with the uniform and laplace distributions, the robustness and power of the F test are quite good. However when the underlying distribution has long tails, such as with the lognormal and cauchy distributions, then the F test becomes conservative, and the power suffers considerably.

Table 2 shows how the actual level of significance varies with the distributions and with the type of test used. In addition to showing the lack of robustness of the F test for long tailed distributions, it shows that the rank transform RT test and the aligned ranks test ART are very stable over the wide range of distributional types studied here. It also shows that the chi-squared version of the Durbin test D1 is consistently conservative, and the F version of the Durbin test D2 is consistently liberal. Not shown in the table is the fact that D2 varies widely from design to design over all of the distributions, ranging from lows of 0.000 to highs of over 11%, underscoring the instability of the F approximation for the F statistic computed on the Durbin ranks.

Table 2. Percentage of times the null hypothesis was rejected, averaged over the 13 BIB designs of Table 1, for random variables with various distribution functions.

<u>Distribution</u>	<u>RT</u>	<u>F</u>	<u>ART</u>	<u>D1</u>	<u>D2</u>
Normal	.0518	.0484	.0539	.0249	.0577
Lognormal	.0518	.0338	.0520	.0249	.0577
Laplace	.0529	.0440	.0554	.0264	.0612
Uniform	.0532	.0529	.0557	.0258	.0600
Cauchy	.0530	.0171	.0537	.0255	.0604

Another way to look at the goodness of the F distribution as an approximation for small and moderate sample sizes, is with a Chi-squared Goodness-of-fit Test, as reported in Table 3. The entire F distribution (or chi-squared distribution in the case of

the Durbin D1 test) was divided into 10 intervals of equal probability, so the expected number of observations in each interval was 250. Then the observed number of observations in each interval, from the 2500 simulations, was compared with the expected number 250 in the usual manner for this goodness-of-fit test.

The chi-squared statistics are given below for normally distributed error terms. Note that values that exceed 16.919, the .95 quantile of the chi-squared distribution with 9 degrees of freedom, are significant at the .05 level, and are denoted by \*. Values that exceed 21.666 are significant at the .01 level and are denoted by \*\*. From these results it is easily seen that the asymptotic approximations are unsatisfactory for both versions of the Durbin test, but are surprisingly good for both the rank transformation procedure RT and the aligned ranks test ART.

Table 3. Values of the chi-squared goodness-of-fit statistic for comparing the F distribution (or chi-squared distribution in the case of D1) to the actual output of 2500 simulations of normally distributed random variables, for 13 BIB designs.

<u>BIB Design</u> txb(k,r, $\lambda$ )	<u>Statistical Test</u>				
	<u>RT</u>	<u>F</u>	<u>ART</u>	<u>D1</u>	<u>D2</u>
1. 4x6(2,3,1)	15.1	10.0	25.6**	6160**	6160**
2. 4x4(3,3,2)	17.6*	7.4	16.7	1505**	1482**
3. 5x10(2,4,1)	19.2*	12.0	35.5**	2929**	2929**
4. 5x5(4,4,3)	25.1**	19.8*	27.1**	263**	92.1**
5. 5x10(3,6,3)	13.0	12.4	10.2	328**	149**
6. 6x10(3,5,2)	4.9	8.2	14.0	132**	194**
7. 6x6(5,5,4)	5.9	9.1	2.6	234**	21.2*
8. 7x7(3,3,1)	6.8	7.2	12.4	384**	403**
9. 7x7(4,4,2)	11.3	14.8	8.4	187**	12.3
10. 7x7(6,6,5)	19.3*	6.4	19.2*	42.8**	23.2**
11. 8x8(7,7,6)	8.5	8.6	6.2	16.6	9.1
12. 9x9(8,8,7)	7.6	16.7	3.5	197	8.8
13. 10x10(9,9,8)	9.9	7.9	5.4	145	6.8

Simple averages of the chi-squared goodness-of-fit test statistics over the 13 BIB designs for the normal distribution, given above, and in addition for lognormal, laplace, uniform, and cauchy distributions averaged over the same 13 designs, are given in Table 4.

Table 4. Chi-squared goodness-of-fit statistics, averaged over the 13 BIB designs of Table 3, for random variables with various distribution functions.

<u>BIB Design</u> txb(k,r, $\lambda$ )	<u>Statistical Test</u>				
	<u>RT</u>	<u>F</u>	<u>ART</u>	<u>D1</u>	<u>D2</u>
Normal (above)	12.6	10.8	14.4	963.3	883.8
Lognormal	12.6	119.4	13.1	963.4	883.9
Laplace	10.9	13.4	15.5	914.3	839.3
Uniform	11.2	8.8	14.9	954.7	864.9
Cauchy	10.3	719.6	14.1	961.5	860.1

In summary, the F distribution provides a good approximation to the null distribution of the rank transformation statistic RT in all of the cases studied. Also the  $\alpha$  level is reasonably accurate.

For the parametric F test, the F approximation is good with the normal distribution, for which it is derived, but also for the laplace and uniform distributions, showing the robust nature of the F test with short tailed distributions. With long tailed distributions such as the lognormal and cauchy, the F test is not robust, and the  $\alpha$  levels are quite conservative.

The aligned ranks test ART behaves in a satisfactory manner for all five of the distributions studied. The fit to the F approximation is uniformly good, but not as good as the rank transformation test. Also the  $\alpha$  levels are close to the target value .05 but again not as close as the rank transformation test.

Both versions of the Durbin test, D1 which uses the chi-squared distribution as the approximation and D2 which uses the F distribution as the approximation, clearly are unsatisfactory both with regard to the asymptotic approximations and with regard to the  $\alpha$  levels. The D1 test is very conservative, while the  $\alpha$  level of the D2 test has erratic behavior depending on the choice of BIB design.

3. THE POWER OF THE RANK TRANSFORMATION TEST. The power of the rank transformation test RT was compared with its most obvious competitors F, ART, D1, D2, and the version D3 of the Durbin test that uses the exact distribution of the Durbin test statistic (Van der Laan and Prakken, 1972). This latter test was included because the poor approximations available for both D1 and D2 made a study of the power of those tests highly unreliable.

Only the BIB designs #6 and #11 were selected for the power study. BIB design #6 has only 30 observations, but is one in which both the RT and the ART have  $\alpha$  levels close to .05. BIB design #11 has 56 observations, which is closer to the large-sample case than #6, and also has well-behaved  $\alpha$  levels for the RT and the ART tests. The only distributions studied were the normal distribution, to see how these methods compared with the F test which is optimal for this case, the lognormal distribution as a representative of long-tailed distributions, and the laplace distribution as a representative of a non-normal short-tailed distribution. The percentage of rejections out of 2500 replications was recorded, both for small treatment effects and for large treatment effects. The results appear in Table 5.

The power simulation results in Table 5 show that the rank transformation test has almost the same power as the F test when the distributions are normal, slightly more power than the F test when the distributions are laplace, and considerably more power than the F test when the distributions are lognormal. The aligned ranks test ART also has almost as much power as the F test when the distributions are normal, and more power than the F test when the



distributions are laplace or lognormal, but not as much power as the rank transformation test in these latter situations. The exact Durbin test D3 suffers uniformly from a lack of power as compared with the rank transformation test, and in most cases has even less power than the aligned ranks test ART. The chi-squared approximation version of the Durbin test D1 has even lower power, due to the extreme conservative nature of the test. The F approximation version of the Durbin test D2 sometimes shows favorable power, but this is due to artificially high actual  $\alpha$  levels in those cases.

In summary, the robust rank transformation procedure appears to have more power overall than any of its competitors, parametric or nonparametric.

Table 5. Power (percentage of rejections) for five tests, two BIB designs, three distributions, and small or large treatment effects, as estimated from 2500 simulations in each case.

SMALL TREATMENT EFFECTS							
Distribution	BIB Design	RT	F	ART	D3	D1	D2
Normal	#6	.162	.160	.158	.119	.079	.150
	#11	.290	.298	.292	.219	.079	.150
Lognormal	#6	.162	.021	.137	.109	.079	.150
	#11	.284	.058	.226	.225	.219	.266
Laplace	#6	.116	.103	.119	.089	.059	.123
	#11	.228	.165	.201	.146	.160	.208

LARGE TREATMENT EFFECTS							
Distribution	BIB Design	RT	F	ART	D3	D1	D2
Normal	#6	.526	.536	.544	.365	.266	.429
	#11	.916	.922	.915	.840	.266	.429
Lognormal	#6	.529	.095	.405	.372	.266	.429
	#11	.894	.287	.792	.845	.840	.868
Laplace	#6	.372	.302	.344	.248	.179	.307
	#11	.748	.624	.714	.635	.626	.683

The theoretical asymptotic relative efficiency (ARE) for the BIB designs has been shown by the author to follow the same ARE formulas as the ones presented by Hora and Iman (1988) for randomized complete block (RCB) designs. The sufficient conditions include the same conditions presented in Section 2 for the asymptotic F distribution under the null hypothesis, plus the same general conditions used by Hora and Iman (1988) in their paper.

They are not repeated here.

The specific ARE results are a function of several factors.

1. The underlying population distribution.
2. The size of the block effects.
3. The number of treatments.

The reader is invited to see the paper by Hora and Iman (1988) for excellent graphs and discussion. Those same graphs and discussions apply to the BIB designs with the same number of treatments as in the RCB designs. The number of blocks, of course, goes to infinity as a condition for the asymptotic nature of the calculations. Their ARE results parallel the power study results presented in Table 5 of this paper

4. SUMMARY AND ADDITIONAL COMMENTS. The rank transformation test is a valid, powerful alternative to existing parametric and nonparametric tests for analyzing balanced incomplete block designs.

Although the discussion in this paper concentrated on replacing the observations by their ranks in an overall ranking, scores based on ranks can be used in the analysis as well, for scores  $a_{ij}$ , given by the usual equations

$$a_{ij} = \phi(E[U_{bk}^{(R_{ij})}]) = \phi\left(\frac{R_{ij}}{bk + 1}\right)$$

and

$$a_{ij} = E[\phi(U_{bk}^{(R_{ij})})]$$

where  $U_{bk}^{(R_{ij})}$  is the  $R_{ij}$  th order statistic from a uniform (0,1) sample of size  $bk$ . Sufficient conditions for the asymptotic results to hold the same as for ranks are for  $\phi(u)$  to be non-constant over its range (0,1), and for the first derivative  $\phi'(u)$  to be absolutely continuous and bounded on (0,1). This latter condition can be relaxed enough to include normal scores.

At this time no general extensions of these results to general linear models appear to exist. Some special cases of the rank transformation are known not to work, such as the test for interaction in a two-way layout with both treatment effects being present. See Thompson (1991) or Blair, et al. (1987) for a discussion of this limitation.

## REFERENCES

- Blair, R.C., Sawilowsky, S.S., and Higgins, J.J. (1987). Limitations of the Rank Transform Statistic in Tests for Interaction. Communications in Statistics - Simulation, 16(4), 1133-1144.
- Cochran, W.G. and Cox, G.M. (1957). Experimental Designs, 2nd Ed.. John Wiley & Sons, New York.
- Conover, W.J. (1980). Practical Nonparametric Statistics, 2nd Ed.. John Wiley & Sons, New York.
- Conover, W.J. and Iman, R.L. (1981). Rank Transformations as a Bridge Between Parametric and Nonparametric Statistics. The American Statistician, 35, 124-128.
- Thompson, G.L. (1991). A Note on the Rank Transformation for Interactions. Biometrika, 78 (3), 697-701.
- Van der Laan, P. and Prakken, J. (1972). Exact Distribution of Durbin's Distribution-Free Test Statistic for Balanced Incomplete Block Designs, and Comparisons with the Chi-Square and F-Approximation. Statistica Neerlandica, 26, 155-164.
- Yates, F. (1936). Incomplete Randomized Blocks. Annals of Eugenics, 7, 121-140.

## THE APPLICATION OF META-ANALYSIS TO ARMY ISSUES

Carl B. Bates and Franklin E. Womack  
U.S. Army Concepts Analysis Agency  
Bethesda, Maryland 20184-2797

**ABSTRACT.** The Army has been studying the employment of scout helicopters as members of attack helicopter teams for over 25 years. The mission of the scout helicopter is to acquire and identify targets and to hand off targets to and coordinate movement of the attack helicopters on the team. It is hypothesized that the effectiveness of the team is increased by the inclusion of scout helicopter(s) on the team. Very few experiments have been designed to test this hypothesis directly. However, it was felt that much data must have been generated on this subject as parts of other field and computer simulated experiments. The Math Stat Team at the U.S. Army Concepts Analysis Agency collected data from many of these previous experiments and undertook to apply meta-analysis techniques to extract information on scout helicopter contribution. This paper discusses the approach taken to the scout helicopter question and touches on possible problems in applying meta-analysis to similar Army issues.

1. **PROBLEM.** The problem is to apply and assess the application of meta-analysis to scout helicopter effectiveness data.

2. **BACKGROUND.** A review article on meta-analysis by Mann in the August 1990 issue of Science stimulated thinking at the US Army Concepts Analysis Agency (CAA). It was conjectured that meta-analysis may have applicability to land combat issues. The Director, CAA, directed that an in-house assessment be made of the applicability of meta-analysis to Army issues. The scout helicopter effectiveness issue was selected because it has been the subject of extensive study during the last three decades. Therefore, a suitably large body of documented work would exist to test meta-analytic techniques.

3. **SCOPE.** The scope of the analysis was limited to data in reports on studies conducted since 1960 on the scout helicopter effectiveness.

### 4. ASSUMPTIONS

a. All important studies involving scout helicopter effectiveness have been submitted to the Defense Technical Information Center (DTIC).

b. A common hypothesis can be extracted from the studies.

c. Some studies will contain data that can be employed in the meta-analysis.

5. **METHODOLOGY.** A schematic of the methodology employed is shown in Figure 1. A literature search was conducted of DTIC documents and the catalog of US Army Combat Developments Experimentation Command (USACDEC) experimentation. A bibliography was then prepared of scout helicopter reports that would be searched for data. Concurrent with the development of the scout helicopter bibliography, meta-analysis literature was researched and studied. The bibliography was studied and a common hypothesis was developed. Appropriate

data were then extracted where available from the scout reports. Meta-analytic methods which would be employed were selected. The selected methods were then applied to the extracted data, and an assessment was made of the application.

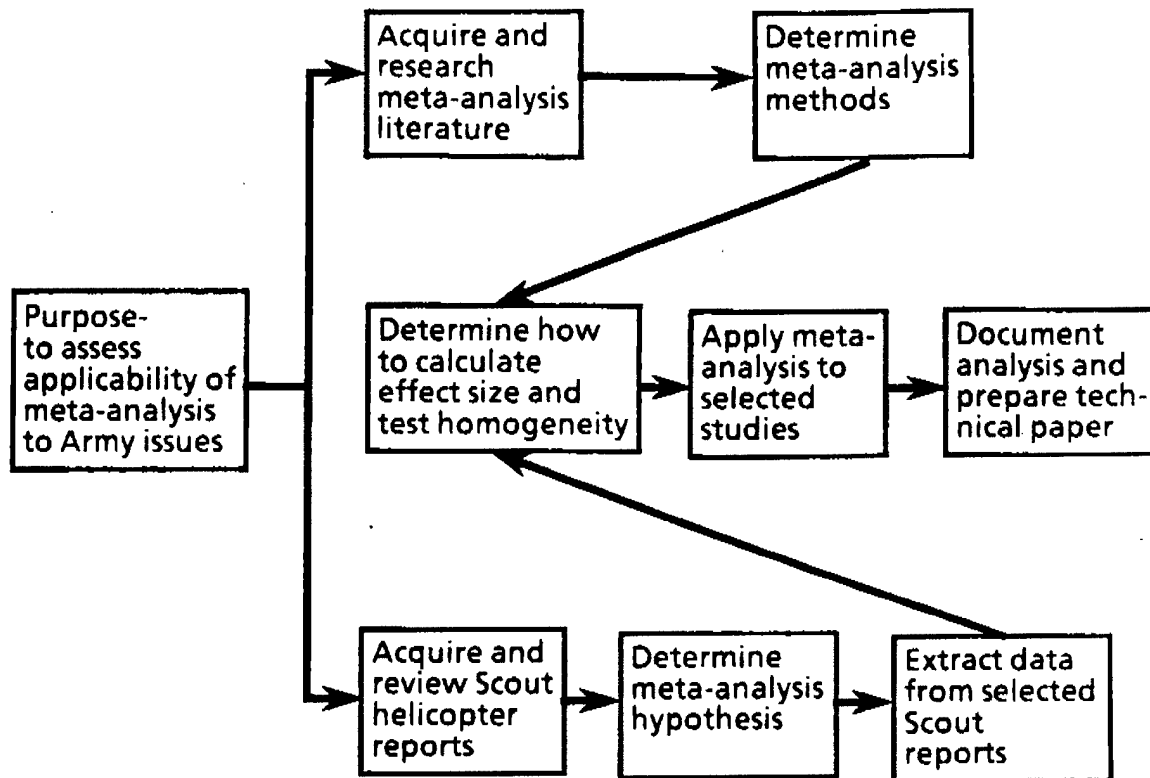


Figure 1. Methodology

## 6. ESSENTIAL ELEMENTS OF ANALYSIS

a. What common hypothesis can be developed from the collective helicopter studies? The null hypothesis is--the scout helicopter did not enhance effectiveness; the alternative hypothesis is--the scout helicopter did enhance effectiveness.

b. What are the underlying assumptions for valid application of meta-analysis?

- (1) Similarity of experiments
- (2) Independence of experiments
- (3) Common unit of measurement

- (4) Absence of covariates affecting experiments
- (5) Homogeneity of effect size, a measure of association between treatment and response
- (6) Individual experiments are summarized and analyzed in a common manner
- (7) All relevant experiments are included

c. What are the major criticisms of meta-analysis? Meta-analysis is controversial. The proponent school appears to be those in the social sciences and the medical field who are searching for ways to improve the review and synthesis of research analysis from separate studies. The opponents appear to be statisticians critical of the statistical methods employed. The misapplication of methods invalidates the statistical analyses. The major criticisms of meta-analysis are:

- (1) Logical conclusions cannot be drawn by comparing and aggregating dissimilar studies that include different measuring techniques, definitions of variables (e.g., treatments, outcomes), and subjects.
- (2) Results of meta-analyses are uninterpretable because results from "poorly" designed studies are included along with results from "good" studies.
- (3) Published research is biased in favor of significant findings because nonsignificant findings are rarely published; this, in turn, leads to biased meta-analysis results.
- (4) Multiple results from the same study are often used which may bias or invalidate the meta-analysis and make the results appear more reliable than they really are, since these results are not independent.

d. What is the assessment of the application of meta-analytic techniques to scout helicopter effectiveness?

(1) Scout helicopter studies are written for a specific purpose. Many study reports do not contain data needed to perform a meta-analysis. Each study addresses a different question and involves different postures, environments, scenarios, and modes of operation. Outcome measures employed in the studies vary. Data sets are not independent.

(2) Meta-analytic methods are not applicable for confirmatory analysis of data from studies of land combat issues that exhibit the characteristics of the scout helicopter data. However, meta-analytic techniques are applicable for exploratory data analysis (EDA) purposes.

7. META-ANALYTIC METHODS. Meta-analytic methods were developed to satisfy a need for combining test results from independent tests in which the same null hypothesis was tested. The tests were designed and conducted for the purpose of testing the same null hypothesis. Methods discussed by Hedges and Olkin (1985) are mostly applicable for measurement data. Rosenthal (1984) and Wolf

(1986) each present a multiplicity of meta-analytic procedures. The major ones are:

**a. Vote-counting.** A common method of combining research results is the so-called "vote-counting" method. Study results are classified into three mutually exclusive categories. The relationship between the independent and dependent variable is either significantly positive, significantly negative, or there is no significant relationship in either direction. Hedges and Olkin (1980) show that the method may tend to make the wrong decision more often as the amount of evidence increases. If the average power of the statistical tests is smaller than the cutoff criterion, the probability that the vote count makes the correct decision tends to zero as the number of studies increases. Moreover, tallies of statistical significance or nonsignificance tell little about the strength or importance of a relationship, Glass (1977). Consequently, the vote-counting method is not recommended.

**b. Combining Contingency Tables.** Another method for synthesizing study results that might appear natural is pooling the raw data. Suppose Study A (Table 1) gave the following categorization of subjects, Glass (1977).

Table 1. Study A

	Treatment	Control	
Improved	50	30	80
Not improved	60	40	100
	110	70	180

The improvement rate of the Treatment (50/110) over the Control (30/70) is 0.45 versus 0.43. Study B gave the following results (Table 2). Here the improvement rate of the Treatment (60/90) is 0.67 versus 0.64 for the Control (90/140).

Table 2. Study B

	Treatment	Control	
Improved	60	90	150
Not improved	30	50	80
	90	140	230

Pooling the two studies gives Table 3. Now the improvement rate for the Treatment (110/200) is 0.55, and the improvement rate for the Control (120/210) is 0.57.

Table 3. Studies A and B

	Treatment	Control	
Improved	110	120	230
Not improved	90	90	180
	200	210	410

Each study showed the Treatment was better than the Control, but the aggregate of the two studies showed the Treatment to be worse than the Control. This so-called Simpson's paradox illustrates why raw data should not be pooled. Rosenthal (1984) gives more dramatic examples of Simpson's paradox. The problem has nothing to do with statistical inference. The problem lies in the unbalanced experimental designs.

c. Fisher's Test. As stated above, many procedures have been proposed for combining probabilities. Virtually any test statistic may be used, converted to p-values, and employed to perform a meta-analysis. A popular procedure was proposed by Fisher (1932). Given  $k$  independent tests or studies which were designed to test the same hypothesis, determine the p-value from each of the  $k$  tests. If the test statistic is from a continuous distribution, the p-values are uniformly distributed. Then,

$$-2 \sum_{i=1}^k \ln p_i \sim \chi^2(2k).$$

$\chi^2(2k)$  provides an overall omnibus test of the common null hypothesis against the common alternative test. If  $\chi^2(2k) \geq \chi^2(2k, 1-\alpha)$ , reject the null hypothesis at the  $\alpha$ -level of significance; otherwise, do not reject the null hypothesis. Mosteller and Bush (1954) found that in situations in which most studies showed results in one direction with p-values close to 0.5, the test would give overly conservative results. However, Fisher's test has been shown to be more asymptotically optimal than other combined tests.

d. Wallis' Test. A statistical test used on the single experiment in this paper is Fisher's exact test. The test statistic for this test has a hypergeometric distribution. This is a discrete distribution. Discrete test statistics do not yield p-values with a uniform distribution. Therefore, Fisher's combined method is inappropriate. If Fisher's method is used it leads to underestimates of significance.

The exact test to evaluate the p-value from combined p-values derived from independent single experiments having discrete test statistics is Wallis'



test (1942). The combined p-value is the probability of realizing a product of p-values as small as or smaller than the product of the p-values taken over all of the single experiments. One first develops the multivariate discrete distribution of the test statistics. Associated with each sample point is a product of p-values. A probability mapping is made from the multivariate sample space of test statistics to the univariate sample space of possible products of p-values. From this transformed space the p-value for the combined experiments is determined by cumulating the probability in this univariate space for values of products less than or equal to the product of p-values for the test statistics observed in each of the single experiments. Computations for all but the simplest cases render this method impractical to implement.

**e. Lancaster's Approximation.** Lancaster (1949) developed a modification to Fisher's combined test. Two sets of probabilities for each single experiment are evaluated. The first probability is the usual p-value. A second probability expressing the probability of observing a test statistic more extreme than the observed value is also determined for each single experiment. A so-called mean value chi-square statistic is determined from the mean of the inverse distribution function. The sum of these derived chi-square statistics over all of the individual experiments is approximately chi-square distributed. However, the approximation tends to overestimate significance in some instances.

**f. Sethuraman's Sequential Test.** Sethuraman's sequential test (1991) is a three-step test based on Fisher's combined test. A significance level is chosen (i.e., say 0.05). The first step is to apply Fisher's combined test. Reject the null hypothesis if the resulting p-value is 0.05 or less. If the p-value at the first step is greater than 0.05, proceed with the second step. Step two consists of adjusting the treatment successes in the contingency table for each single experiment by adding one. The other cells are adjusted so that the marginal totals remain the same as in step one. Fisher's combined test is applied again. If the p-value at step two is greater than 0.05, there is not sufficient information to reject the null hypothesis and the test concludes at step two. If the p-value at step two is less than 0.05, testing continues to step three. At step three, a uniform random number (0, 1) is selected for each single experiment. A weighted p-value is determined for each single experiment by multiplying the p-value obtained for the original contingency table by the random number selected for the experiment and multiplying the p-value obtained in the adjusted table of step two by one minus the same random number. Fisher's combined test is applied to this set of weighted p-values. The null hypothesis is rejected if the combined p-value is less than 0.05. Otherwise, there is not sufficient information to reject the null hypothesis.

**g. Mantel-Haenszel-Peto (MHP) Method.** Mantel and Haenszel (1959) developed a method for combining rates from clinical trials. Richard Peto, Oxford University, modified the procedure. The MHP method appears to be the most popular meta-analytic technique used for synthesizing clinical trial study results. The conventional layout for clinical trials is a 2x2 contingency table as shown in Table 4, where a, b, c, and d are the frequencies of the four mutually exclusive outcomes.

Table 4. Clinical Trial

	Treatment	Control	
Improved	a	b	a + b
Not improved	c	d	c + d
	a + c	b + d	N = a + b + c + d

The odds ratio (also called the cross-product ratio) is the ratio of the odds associated with the Treatment to the odds associated with the Control. The odds ratio is simply  $ad/bc$ . The observed number ( $O$ ) improved after receiving the Treatment is  $a$ . The expected number ( $E$ ) is  $(a+c)(a+b)/N$ . Under the null hypothesis of no Treatment effect,  $O-E$  should vary about zero with variance  $V = E[(N-(a+c))/N][(N-(a+b))/(N-1)]$ . Under the null hypothesis of no Treatment effect, the statistic

$$\sum_{i=1}^k (O_i - E_i)^2 / \sum_{i=1}^k V_i$$

is approximately chi-squared distributed with 1 degree of freedom, and  $k$  is the number of independent clinical trials. If the Treatment is beneficial,  $O-E$  tends to be positive; if there is no difference,  $O-E$  is zero. Yusuf et al. (1985) and Berlin et al. (1989) discuss the test.

**h. Der Simonian-Laird-Cochran Method (DLC).** The DLC (1986) is another method for combining data from similar single experiment contingency tables. Like coin tossing experiments, it is assumed that a proportion of favorable events exists for the treatment and the control groups. The interest lies in the proportional difference between the two groups. In addition, the DLC attempts to measure an among experiment variation component much as the random effects ANOVA model. Weights for each experiment are determined by the inverses of a combination of within and among experiment variation.

**i. Logistic Regression.** Logistic regression is a generalization of the MHP method. In addition to a primary predictor variable (i.e., Scout or No Scout) used in MHP and DLC, logistic regression allows one to include other predictor variables. Usually these other variables would be other covariates which are not otherwise controlled for or differ from single experiment to single experiment. Logistic regression is based upon the binomial distribution as opposed to regression which is usually based on the normal distribution. Being able to include covariates in the analysis helps in two ways. First, if there are any important covariates which are uncontrolled for, they would contribute to the heterogeneity of combined experiments and lead to an inflated variance. Logistic regression would help by adjusting for this systematic variation and give a more precise measure of the random variation. Second, it allows one to measure the effect of significant interactions between the primary predictor variable and the other covariates.

**8. DATA BASE.** One hundred twenty-eight scout helicopter reports were acquired and reviewed. Twelve reports contained data which might be extracted to use in the meta-analysis. One hundred thirty-nine separate experiments provided count data which could be assembled into 2x2 contingency tables similar to that illustrated in Table 4. These 12 reports contained data representing seven different groups of measures of effectiveness and came from five different testing environments. The 139 experiments are cross-tabulated in Table 5 by measure of effectiveness block, study report number (i.e., a number from 01 to 128 representing the order of a report's acquisition), and testing environment (i.e., operational testing or one of four different simulation models).

**Note:** A two-letter acronym is used for each of the seven measures of effectiveness as follows: (1) BK - kills by all Blue weapons engaged; (2) BS - survivability of all Blue weapons; (3) HK - kills by Blue helicopters; (4) HS - survivability of Blue helicopters; (5) DD - detection; (6) EE - engagement, and (7) SE - subjective evaluations.

**Table 5. Cross-tabulation of Experiments by Group, Study, and Model**

Report Block	Operational testing					Model CARMONETTE						Other models			Block total
	01	07	16	27	28	14	20	31	33	49	57	AVBATS 17	AVWAR 20	JANUS 49	
BK							3	5	8	3	6	1			26
BS							3	5	8	3	6	1			26
HK			1	2		1		5	8	3	6	1		3	30
HS				2			3	4	8	3	6	2	1	3	32
DD	1	12	1	1	1								1		17
EE				5	1										6
SE			2												2
Report total	1	12	4	10	2	1	9	19	32	12	24	5	2	6	139

**9. ANALYSIS.** One parameter of interest is the proportional difference between the treatment and the control. If the scout enhanced effectiveness, a measure of this for each experiment would be the difference between the proportion of the trials showing improvement when the scout was employed in the experiment versus the proportion of the trials showing improvement when the scout was not employed. This difference can be calculated for each experiment where the counts are tabulated like the format presented in Table 4. This difference is  $a/(a + c) - b/(b + d)$ . The null hypothesis is that scout does not enhance effectiveness. If this is true, we would expect the proportional difference to be around zero in case the scout is equally as effective as without the scout or a negative difference in case the without scout case is more effective than with scout. On the other hand, if the scout enhances effectiveness, we would like to reject the null hypothesis in favor of the alternative hypothesis which states that scout employment enhances effectiveness. In this case, the proportional difference should be significantly positive. When we accumulate the proportional differences from a number of separate experiments, we would expect that there would be only a small amount of variability in these values solely due to random variation. Figure 2 is a histogram of the proportional differences between the scout (treatment) and no scout (control) for each of the 139 experiments considered in this study.

a. Two points relevant to any meta-analysis of the data can be seen in this histogram. First, the data points are very disperse. In fact, the range is 135 percentage points out of the possible spread of 200 points (i.e., from -100 percent to +100 percent). Any estimate of the true parameter, proportional difference of effectiveness between scout and no scout is not very precise. Secondly, a rough estimate of the true parameter is the median of the histogram. The median proportional difference between scout and no scout for the 139 experiments is zero. This is consistent with the null hypothesis of no difference in the effectiveness of with scout helicopters and the effectiveness without scout helicopters. There are several notable outlier experiments. In the spirit of exploratory statistics, it would be interesting to explore why the results of these experiments were so much different than the majority of the other experiments.

b. Our study looked at more sophisticated methods of meta-analysis including (1) Fisher's exact test and several variants (i.e., Wallis' test, Lancaster's approximation, and Sethuraman's sequential test), (2) Mantel-Haenszel-Peto method, (3) DerSimonian-Laird-Cochran method, and (4) logistic regression. It is invalid to apply the methods of meta-analysis to this data base without ignoring one or more of the assumptions made for a valid application. Moreover, when one ignores these assumptions and applies the more sophisticated methods of meta-analysis as we did in our study, no more relevant facts are elicited than are discernible from Figure 2.

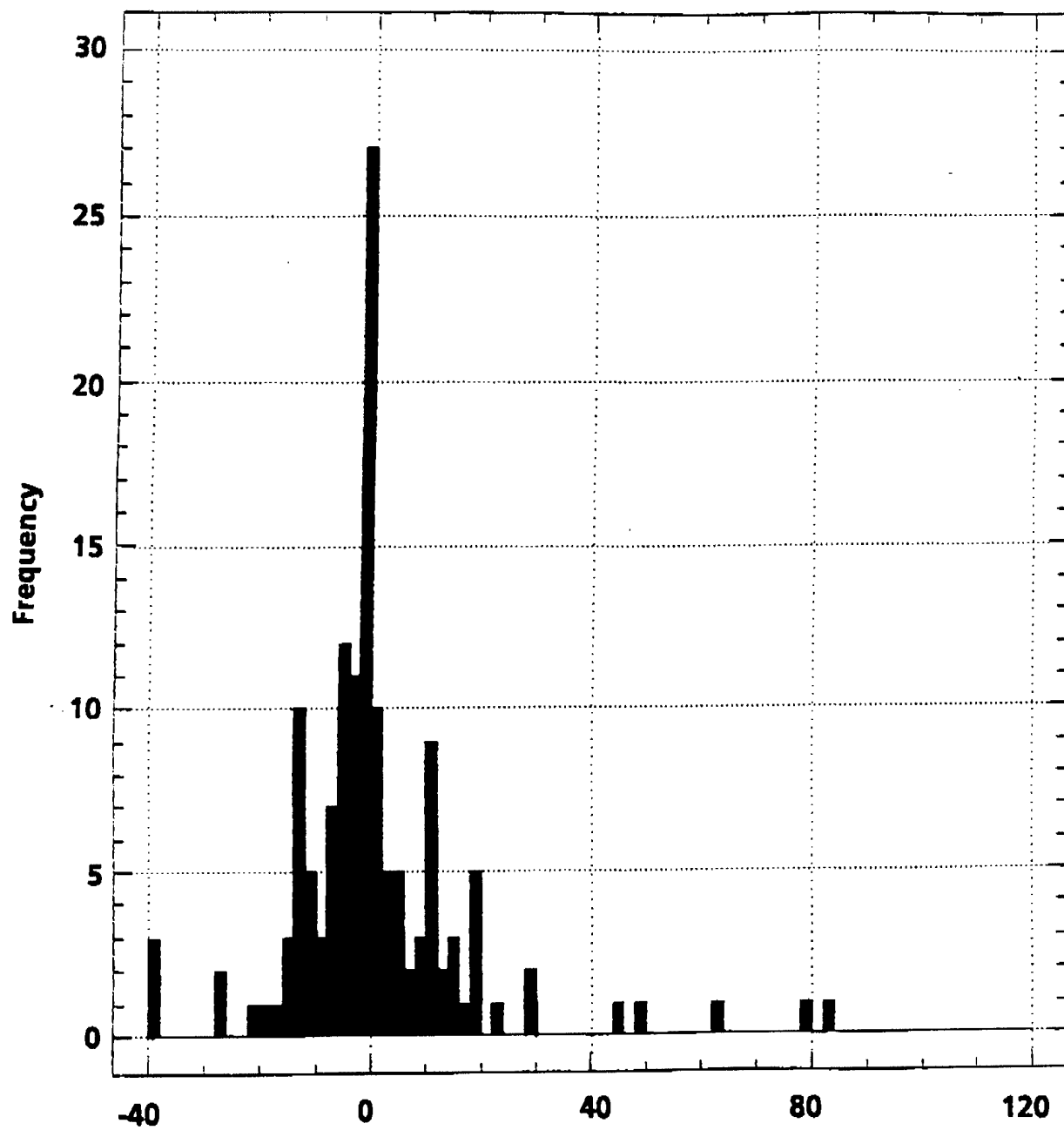


Figure 2. Histogram of Percent Difference - All Experiments

## 10. FINDINGS

- a. Each study addressed a different problem. The purpose, scope, objective, and control conditions varied across studies.
- b. Potentially useful studies could not be included in the analysis because pertinent data were not preserved in the study documentation.
- c. Meta-analytic techniques are not appropriate for the analysis of specific Army data which are heterogeneous as are the scout helicopter data.

## REFERENCES

1. DerSimonian, R. and Laird, N. (1986), "Meta-Analysis in Clinical Trials," Controlled Clinical Trials, Vol.7, pp 177-188, Elsevier Science Publishing Co., Inc., New York, NY
2. Fisher, R. A. (1932), Statistical Methods for Research Workers, 4th Edition, Oliver & Boyd, Ltd., London, England
3. Glass, G. V. (1977), "Integrating Findings: The Meta-Analysis of Research," Review of Research in Education, Vol. 5, pp 351-379
4. Hedges, L. V. and Olkin, I. (1980), "Vote-Counting Methods in Research Synthesis," Psychological Bulletin, Vol. 88, pp 359-369
5. Hedges, L. V. and Olkin, I. (1985), Statistical Methods for Meta-Analysis, Academic Press, Inc., San Diego, CA
6. Lancaster, H. O. (1949), "The Combination of Probabilities Arising from Data in Discrete Distributions," Biometrika, Vol. 36, pp 370-382
7. Mann, C. (1990), "Meta-Analysis in the Breech," Science, Issue 249, pp 476-480
8. Mantel, N. and Haenszel, W. (1959), "Statistical Aspects of the Analysis of Data from Retrospective Studies of Disease," Journal of the National Cancer Institute, Vol. 22, pp 719-748
9. Mosteller, F. and Bush, R. R. (1954), Selected Quantitative Techniques in the Handbook of Social Psychology, Vol. I, pp 289-334, Addison-Wesley Publishing Co., Cambridge, MA
10. Rosenthal, R. (1984), Meta-Analytic Procedures for Social Research, Sage Publications, Inc., Newbury Park, CA
11. Sethuraman, J. (1991), Written Correspondence, Department of Statistics, Florida State University, Tallahassee, FL
12. Wallis, W. A. (1942), "Compounding Probabilities from Independent Significance Tests," Econometrica, Vol. 10, pp 229-248

13. Wolf, F. M. (1986), Meta-Analysis: Quantitative Methods for Research Synthesis, Sage Publications, Inc., Newbury Park, CA
14. Yusuf, S., Peto, R., Lewis, J., Collins, R. and Sleight, P. (1985), "Beta Blockage During and After Myocardial Infarction: An Overview of the Randomized Trials," Progress in Cardiovascular Diseases, Vol. 27, pp 355-371
15. Force Development CATALOG of USACDEC EXPERIMENTATION, (31 Jul 85), TRADOC Pamphlet No. 71-5, Headquarters US Army Training and Doctrine Command, Fort Monroe, VA
16. Publishing Record of CDEC/TEC Test Reports (09 Oct 90)
17. An Application of Meta-analysis, CAA-TP-91-5, US Army Concepts Analysis Agency, Sep 91 (CONFIDENTIAL)

# Total Time on Test Function Orthogonal Components and Tests of Exponentiality

W. D. Kaigh and Alexander K. White

## ABSTRACT

Mathematical development reminiscent of Fourier analysis applied to the sample total time on test function (TTT) yields scale-free orthogonal components analogous to empirical quantile function (EQF) component  $L$ -statistics utilized by Kaigh (1992a) for assessing one-sample uniformity. As estimators of TTT Fourier coefficients with respect to the complete orthonormal set of Legendre polynomials, the TTT components are linear combinations of normalized spacings with Hahn polynomial vector weight functions to provide directional criteria for assessment of departures from exponentiality. In particular, the first TTT component is equivalent to the cumulative total time on test statistic and Gini statistic investigated by Gail and Gastwirth (1978a).

Analogous to the quadratic smooth tests for exponentiality proposed by Rayner and Best (1986, 1989), aggregates of TTT component squares yield component decompositions of the squared coefficient of variation and a discrete Anderson-Darling type statistic. A simple average of the discrete and conventional Anderson-Darling statistics produces a hybrid exponentiality criterion which exhibits strong performance against various alternatives.

Monte Carlo results indicate adequacy of asymptotics for small samples and empirical power comparisons show that TTT component exponentiality criteria are quite competitive for various alternative models, including those with "bathtub" hazard rates.

Use of the TTT components and omnibus statistics is illustrated by application to real data. A smoothed continuous TTT plot and simple four-number summary are developed for description and presentation purposes.

**Keywords:** Goodness of fit,  $L$ -moment,  $Q$ -statistic, Quantile, Spacing, Uniformity.

**Author's footnote:** W. D. Kaigh is Professor, Department of Mathematical Sciences, University of Texas at El Paso, El Paso, TX 79968. Alexander K. White is a graduate student in the Department of Probability and Statistics, Michigan State University, East Lansing, MI, 48824. This work was performed under sponsorship of the U.S. Army under Contract No. DAAL03-89-G-0098.



## 1. INTRODUCTION AND PRELIMINARIES

Among the many tests available for the composite exponential null hypothesis, several utilize normalized sample spacings and the sample total time on test function (TTT). Most of these, however, employ relatively simple summary measures with perhaps some loss of TTT information. Reminiscent of Fourier analytic methods, more detailed TTT orthogonal component criteria proposed here are based on empirical quantile function (EQF) techniques developed in Kaigh (1992a,b) for the one-sample uniformity and nonparametric two-sample problems. The new omnibus and directional scale-free TTT tests for exponentiality emerge naturally from EQF application of the general approach established primarily for the empirical distribution function (EDF) by Durbin and Knott (1972). The EQF direction pursued here with TTT is opposite that of Rayner and Best (1986, 1989) who focus on the EDF to obtain different orthogonal components and aggregate smooth tests for exponentiality.

With preliminary discussion complete, the remainder of the paper is organized as follows. To provide conceptual insight and establish notation, certain TTT functionals are developed initially. These functionals then lead naturally to the Gini statistic, the squared coefficient of variation, and a new EQF Anderson-Darling type TTT quadratic statistic. In Section 2 individual TTT components are described and orthogonal component decompositions are developed. Monte Carlo power comparisons are presented and discussed in Section 3. An actual data analysis illustration in Section 4 employs significance testing and introduces general numerical and graphical TTT descriptive methods. Brief concluding remarks are presented in Section 5.

### 1.1 TTT Functionals

The main problem under consideration here is the hypothesis that a continuous life distribution with cdf  $F$  and qf  $Q$  is exponentially distributed with cdf  $F_0(x) = 1 - e^{-x/\beta}$ ,  $x > 0$ , and quantile function (qf)  $Q_0(u) = -\beta \log(1-u)$ ,  $0 < u < 1$ , for unspecified mean  $\beta > 0$ . The proposed new exponentiality test criteria, as well as many conventional procedures, employ the total time on test concept. Although not explicitly stated in the following informal treatment of TTT functionals designed to motivate test statistic analogues, necessary differentiability and integrability conditions are assumed to be satisfied whenever appropriate.

Corresponding to the cdf  $F$  and qf  $Q$  with assumed finite mean  $\mu$ , pdf  $f$ , qdf  $q$ , and hazard rate

$\lambda$ , important total time on test functionals are defined on  $[0,1]$  by

$$\begin{aligned}
H^{-1}(t) &= \int_{0 \leq x \leq Q(t)} [1-F(x)] dx && \text{total time on test transform} \\
&= \int_{0 \leq u \leq t} (1-u)q(u) du \\
H^{-1}(t)/H^{-1}(1) &= (1/\mu) \int_{0 \leq x \leq Q(t)} [1-F(x)] dx && \text{scaled total time on test transform} \\
&= (1/\mu) \int_{0 \leq u \leq t} (1-u)q(u) du \\
V_F &= (1/\mu) \int_{0 \leq t \leq 1} H^{-1}(t) dt && \text{cumulative total time on test} \\
&= 1 - (1/\mu) \int_{0 \leq u \leq 1} u(1-u)q(u) du .
\end{aligned}$$

Omitting the usual population Lorenz curve definition of the Gini index  $G_F$ , we merely note that  $G_F = 1 - V_F$ .

General background on the total time on test concept and other related functionals appears in Barlow (1979), Chandra and Singpurwalla (1981), and Shorack and Wellner (1986). In particular,  $H^{-1}$  is concave (convex) if the hazard rate is increasing (decreasing), and the scaled total time on test transform is the identity function for an exponential distribution.

Treatment here will primarily address the TTT derivative  $(d/dt) H^{-1}(t) = (1-t)q(t) = 1/\lambda[Q(t)]$  and its scale-free counterpart. Consider orthogonal representation with respect to the complete orthonormal set of Legendre polynomials  $\{\Pi_k\}_{k \geq 0}$  on the unit interval. The Fourier series representing the TTT derivative is

$$(d/dt) H^{-1}(t) \sim \mu + \sum_{1 \leq k < \infty} \left[ \int_{0 \leq u \leq 1} \Pi_k(u) (1-u)q(u) du \right] \Pi_k(t) . \quad (1)$$

The Legendre polynomials then provide integral Fourier coefficient functionals which characterize the TTT. For exponential distributions with linear TTT, all  $k \geq 1$  Fourier coefficients vanish as a consequence of orthogonality and the fact that  $\Pi_0$  is identically one. Hence, the scaled Fourier coefficients  $(1/\mu) \int_{0 \leq u \leq 1} \Pi_k(u) (1-u)q(u) du$  are easily interpreted measures of departure from the composite null assumption of exponentiality. For example,  $\Pi_1(u) = \sqrt{3}(2u-1)$  provides  $(1/\mu) \int_{0 \leq u \leq 1} \Pi_1(u) (1-u)q(u) du = -\sqrt{12}(V_F - 1/2) = \sqrt{12}(G_F - 1/2)$  with null value zero.

Suggested by the Parseval identity, the TTT derivative scale-free squared norm with value zero

for all exponential distributions is given by

$$\int_{0 \leq t \leq 1} [(1/\mu) (d/dt) H^{-1}(t) - 1]^2 dt = \sum_{1 \leq k < \infty} [(1/\mu) \int_{0 \leq u \leq 1} \Pi_k(u) (1-u)q(u) du]^2. \quad (2)$$

Integration by parts calculations utilizing the Legendre polynomial differential equation  $(d/du)[u(1-u)]\Pi'_s(u) + s(s+1)\Pi_s(u) = 0$  show that the associated Ferrer functions defined by  $\Pi_s^{-1}(u) = [1/s(s+1)]^{1/2} [u(1-u)]^{1/2} \Pi'_s(u)$  are also orthonormal to yield associated inner products

$$\int_{0 \leq t \leq 1} [(1/\mu) H^{-1}(t) - t] [t(1-t)]^{-1} \Pi_s^{-1}(t) dt = (1/\mu) \int_{0 \leq u \leq 1} \Pi_s(u) (1-u)q(u) du. \quad (3)$$

The related functional of primary importance later accumulates weighted TTT deviations from the identity function as

$$\int_{0 < t < 1} [(1/\mu) H^{-1}(t) - t]^2 [t(1-t)]^{-1} dt = \sum_{1 \leq s < \infty} [1/s(s+1)] [(1/\mu) \int_{0 \leq u \leq 1} \Pi_s(u) (1-u)q(u) du]^2. \quad (4)$$

## 1.2. TTT Exponentiality Test Criteria

Various TTT test statistics for the exponential composite null hypothesis are introduced now as sample analogues of the previously defined functionals. Suppose that a random sample  $X_1, \dots, X_n$  with continuous cdf  $F$  produces corresponding order statistics  $0 = X_{0:n} < X_{1:n} < \dots < X_{n:n}$  and normalized sample spacings  $(n-j+1)(X_{j:n} - X_{j-1:n})$ ,  $1 \leq j \leq n$ . As assumed throughout, it is important for subsequent development that zero be a natural minimum of the support of  $F$ .

Observing that normalized spacings preserve the total  $\sum_{1 \leq j \leq n} (n-j+1)(X_{j:n} - X_{j-1:n}) = \sum_{1 \leq i \leq n} X_i$ , form the nondecreasing sample (scaled) TTT by calculating cumulative partial sums

$$S_{j:n} = \sum_{1 \leq i \leq j} (n-i+1) (X_{i:n} - X_{i-1:n}) / \sum_{1 \leq i \leq n} X_i, \quad 1 \leq j \leq n-1 \quad (5)$$

and appending endpoints  $S_{0:n} = 0$ ,  $S_{n:n} = 1$ . Development here focuses on the ordered values  $0 = S_{0:n} < S_{1:n} < \dots < S_{n-1:n} < S_{n:n} = 1$  and their corresponding differences

$$D_j = (n-j+1) (X_{j:n} - X_{j-1:n}) / \sum_{1 \leq i \leq n} X_i, \quad 1 \leq j \leq n. \quad (6)$$

Under the exponential composite null hypothesis, the scale-free statistics defined by (5) and (6) are distributed as order statistics and spacings for a sample of

$n-1$  from the uniform distribution on  $(0,1)$ , respectively, with  $ES_{j:n}=j/n$ ,  $VarS_{j:n}=j(n-j)/n^2(n+1)$ ,  $Cov(S_{i:n}, S_{j:n})=i(n-j)/n^2(n+1)$ ,  $i \leq j$ , and  $ED_j=1/n$ ,  $VarD_j=(n-1)/n^2(n+1)$ ,  $Cov(D_i, D_j)=-1/n^2(n+1)$ ,  $i \neq j$ .

Although various TTT exponentiality tests assess uniformity of values from (5), those of main concern here are the Gini statistic, the squared coefficient of variation, and a discrete EQF Anderson-Darling analogue of (4).

Employing TTT spacings (6), Gail and Gastwirth (1978a) defined the Gini statistic  $G_n$  as

$$G_n = \sum_{1 \leq j \leq n-1} j(n-j) (X_{j+1:n} - X_{j:n}) / (n-1) \sum_{1 \leq i \leq n} X_i. \quad (7)$$

In notation here,  $G_n = \sum_{0 \leq j \leq n-1} j D_j / (n-1)$  and summation by parts shows that (7) is algebraically equivalent to the cumulative total time on test statistic  $V_n = \sum_{1 \leq j \leq n-1} S_{j:n} / (n-1) = 1 - G_n$ .

As the sample analogue of (2), the squared coefficient of variation  $CV_n^2$  for the normalized sample spacings is written as

$$CV_n^2 = [n^2/(n-1)] \sum_{1 \leq j \leq n} (D_j - 1/n)^2. \quad (8)$$

From the TTT spacings moment expressions given previously, it follows that  $CV_n^2$  has null expectation  $1-1/(n+1)$ .

The new TTT omnibus scale-free test criterion for exponentiality under primary consideration here is the discrete analogue of the Anderson-Darling EDF statistic defined by

$$QA_n^2 = n(n+1) \sum_{1 \leq j \leq n-1} (S_{j:n} - j/n)^2 / j(n-j). \quad (9)$$

This quadratic statistic, which is the sample version of (4), was introduced as an EQF test of uniformity in Kaigh (1992b, eq. 2.8). The statistic  $QA_n^2$  has null exponential expectation  $1-1/n$  and limiting distribution that of the asymptotic weighted sum of chi-square variates for the usual Anderson-Darling statistic [see Durbin and Knott (1972) or Shorack and Wellner (1986), p. 225].

Later we show that the Gini statistic is equivalent to the first TTT component in the orthogonal component decomposition of  $QA_n^2$ . Unlike  $G_n$ , which employs only the first TTT component, the

quadratic statistics  $CV_n^2$  and  $QA_n^2$  utilize all sample TTT components. Instead of assigning equal weight to each TTT component, however,  $QA_n^2$  incorporates component damping as suggested by Shorack and Wellner (1986, p. 226) to diminish power dilution effects from nonresponsive high frequency terms. Viewed in this sense, the new criterion  $QA_n^2$  is a compromise between the Gini statistic and the coefficient of variation scale-free tests of exponentiality.

As a basis for comparison with  $QA_n^2$ , the conventional Anderson-Darling statistic contrasts the EDF of the scaled TTT with the identity function on (0,1). For this quadratic criterion, employed here as a test of exponentiality with null mean one, a TTT computational formula (see Shorack and Wellner, 1986 p. 227) is

$$FA_n^2 = (n-1)^{-1} \sum_{1 \leq j \leq n-1} (2j-1) \log[1/S_{j:n} (1-S_{n-j:n})] - (n-1). \quad (10)$$

Incorporating both EDF and EQF contributions, a hybrid exponentiality criterion is defined as the simple average

$$FQA_n^2 = (1/2) (FA_n^2 + QA_n^2). \quad (11)$$

Intuitively appealing, such blending is actually required to avoid a disturbing asymmetry with rank spacings statistics in Kaigh (1992b) for the nonparametric two-sample problem. More general than the one-sample formulation here, a large sample argument demonstrates that the problems are conceptually related.

From orthogonal component decompositions in the next section, it follows that this hybrid statistic also has the same asymptotic null distribution as  $FA_n^2$  and  $QA_n^2$ . For testing the composite hypothesis of exponentiality, asymptotic percentage points for  $FA_n^2$ ,  $QA_n^2$ , and  $FQA_n^2$  at levels .10, .05, .01 are then 1.933, 2.492, 3.857, respectively.

## 2. TTT ORTHOGONAL COMPONENT REPRESENTATIONS

### 2.1 Orthogonal Component Decompositions

Following definition of individual TTT components, decompositions of the TTT quadratic statistics are developed. In contrast with the EDF approach in Rayner and Best (1986, 1989)

employing the continuous Laguerre orthonormal system on  $(0, \infty)$ , we exploit instead the Legendre polynomials on  $[0, 1]$  and the  $\mathbb{R}^n$  orthonormal basis  $\{\pi_{0,n-1}, \pi_{1,n-1}, \dots, \pi_{n-1,n-1}\}$  consisting of Hahn polynomial vectors.

The discrete Hahn polynomial orthonormal vectors are generated by application of the Gram-Schmidt process to the  $n$  vectors with entries of the form  $j^k$ ,  $1 \leq j \leq n$ , for exponents  $k=0, \dots, n-1$ . Related closely to the continuous Legendre polynomials, the Hahn polynomial orthonormal vector basis includes the unit vector of constants  $n^{-1/2} \mathbf{1}_n = n^{-1/2} [1, \dots, 1]^T$  as  $\pi_{0,n-1}$  and higher-order vectors with entries which are linear, quadratic, cubic, etc. [for further background on Hahn polynomials see Kaigh (1992a,b); Neuman and Schonbach (1974); Nikiforov, Suslov, and Uvarov (1991); Rayner and Best (1989)].

Using spacings from (6), the TTT component statistics are (random) inner products defined by

$$Z_{k,n} = -[n(n+1)]^{1/2} \sum_{1 \leq j \leq n} \pi_{k,n-1}(j) D_j, \quad 1 \leq k \leq n-1. \quad (12)$$

Identified later as sample versions of the scaled TTT Fourier coefficients from (1), these scale-free components satisfy  $Z_{k,n} = -(n+1)^{1/2} (T_{k,n} / \bar{X})$  in terms of the normalized spacings inner products

$$T_{k,n} = n^{-1/2} \sum_{1 \leq j \leq n} \pi_{k,n-1}(j) (n-j+1) (X_{j:n} - X_{j-1:n}). \quad (13)$$

For the exponential distribution with mean  $\beta$ , the normalized spacings are iid exponential themselves so  $ET_{k,n} = 0$  and  $\text{Var} T_{k,n} = \beta^2/n$ . Because the exponential sample mean with expectation  $\beta$  and variance  $\beta^2/n$  is a sufficient statistic, Basu's theorem shows that  $Z_{k,n}$  and  $\bar{X}$  are independent. Simple expectation calculations demonstrate that the TTT components (12) are uncorrelated rv's, each with mean zero and variance one under the exponential null hypothesis. Arguments presented later show  $T_{k,n} / \bar{X}$  converges with probability one to the Fourier coefficient of the scaled total time on test functional from (1) for both alternative and exponential distributions.

As an orthonormal basis the Hahn polynomial vectors represent the TTT spacings vector  $\mathbf{D}$  with entries (6) as  $\mathbf{D} = \sum_{0 \leq k \leq n-1} (\pi_{k,n-1}^T \mathbf{D}) \pi_{k,n-1} = \mathbf{1}_n/n + \sum_{1 \leq k \leq n-1} (\pi_{k,n-1}^T \mathbf{D}) \pi_{k,n-1}$ . This relation, which shows that the  $n-1$  components in (12) written as  $Z_{k,n} = -[n(n+1)]^{1/2} \pi_{k,n-1}^T \mathbf{D}$  are

algebraically equivalent to the TTT, establishes a components decomposition analogue of (2) as

$$CV_n^2 = [n/(n-1)(n+1)] \sum_{1 \leq k \leq n-1} Z_{k,n}^2. \quad (14)$$

Demonstrating equivalence of several quadratic exponentiality tests, Currie and Stephens (1986) showed that  $CV_n^2$  is essentially the Greenwood (1946) statistic for uniformity applied to TTT.

To avoid further mathematical digression and notation, a similar components decomposition for  $QA_n^2$  derived along the lines of (3) and (4) (see Kaigh, 1992a) is merely stated here as

$$QA_n^2 = \sum_{1 \leq s \leq n-1} [1/s(s+1)] Z_{s,n}^2. \quad (15)$$

The above representation for  $QA_n^2$  employing EQF components is similar in form to that for  $FA_n^2$  in Shorack and Wellner (1986, p. 225) involving EDF components. Results in Kaigh (1992a, Theorem 2), which demonstrate null asymptotic equivalence of EQF and EDF uniformity components, establish identical limiting distributions for the scale-free statistics  $FA_n^2$ ,  $QA_n^2$ , and  $FQA_n^2$  under exponentiality.

## 2.2 Components and Fourier Coefficient Estimators

Attention now will focus on individual TTT components (12) and their asymptotic distribution theory. Results under the composite null hypothesis are reproduced from Kaigh (1992a) as

**Theorem 1.** Suppose  $X_1, \dots, X_n$  are iid exponential rv's with mean  $\beta$  and TTT spacings vector  $D = [D_1, \dots, D_n]^T$  with entries  $D_j = (n-j+1) (X_{j:n} - X_{j-1:n}) / \sum_{1 \leq i \leq n} X_i$ . The TTT components are random inner products with respect to the Hahn polynomial vector orthonormal basis given by  $Z_{s,n} = -[n(n+1)]^{1/2} \pi_{s,n-1}^T D$ ,  $1 \leq s \leq n-1$ . These components are uncorrelated statistics satisfying

- i)  $E Z_{s,n} = 0$   
 $E Z_{s,n}^2 = 1$   
 $E Z_{s,n}^3 = -2[n^{1/2}(n+1)^{1/2}/(n+2)] \sum_{1 \leq i \leq n} [\pi_{s,n-1}(i)]^3$   
 $E Z_{s,n}^4 = [n(n+1)/(n+2)(n+3)] \{3 + 6 \sum_{1 \leq i \leq n} [\pi_{s,n-1}(i)]^4\}$   
 $E Z_{s,n}^2 Z_{t,n}^2 = [n(n+1)/(n+2)(n+3)] \{1 + 6 \sum_{1 \leq i \leq n} [\pi_{s,n-1}(i)]^2 [\pi_{t,n-1}(i)]^2\}, s \neq t$
- ii)  $Z_{s,n} \Rightarrow N(0,1)$  for each fixed  $s$  as  $n \rightarrow \infty$ .

It can be shown that the null third and fourth moments above converge to the limiting standard normal values and that  $\text{Corr}(Z_{s,n}^2, Z_{t,n}^2) \rightarrow 0$  as  $n \rightarrow \infty$ . Further distinguishing the TTT components here from those in Rayner and Best (1986, 1989), only asymptotically do the corresponding Laguerre EDF components produce zero null means.

Using finite-difference notation  $\Delta g(x) = g(x+1) - g(x)$ , summation by parts shows that the TTT spacings component (9) admits the  $\underline{L}$ -statistic representation

$$Z_{k,n} = [n(n+1)]^{1/2} \sum_{1 \leq j \leq n-1} [\Delta_j \pi_{k,n-1}(j)] (S_{j:n} - j/n).$$

Because the first Hahn polynomial vector  $\pi_{1,n-1}(j) = [12/(n-1)n(n+1)]^{1/2} [j - (n+1)/2]$  is linear, the TTT location component  $Z_{1,n} = [12(n-1)]^{1/2} (V_n - 1/2)$  is equivalent to the cumulative total time on test and Gini statistics, standardized to have null mean zero and variance one. Sharing an asymptotic minimax property with  $V_n$ , the statistic  $G_n$  was employed as a benchmark criterion by Rayner and Best (1986, 1989) for evaluation of their EDF smooth tests for exponentiality. Close connections with the  $\underline{L}$ -moments in Hosking (1990) indicate that the second, third, and fourth order components, respectively, measure TTT scale, skewness, and kurtosis departures from the identity function.

We examine first the Fourier coefficient functionals in (1) and then related behavior of individual TTT components under alternative models. Writing  $\mu_{r:k}(Q)$  for the mean of the  $r$ th order statistic in a sample of size  $k$  from  $F$ , calculations similar to those in Kaigh (1992a, sec. 4.3) show

$$\int_{0 \leq t \leq 1} \Pi_k(t)(1-t)q(t) dt = (2k+1)^{1/2} (k+1)^{-1} \sum_{0 \leq r \leq k} (-1)^{k-r} \binom{k}{r} (k-r+1) [\mu_{r+1:k+1}(Q) - \mu_{r:k+1}(Q)].$$

The  $\underline{U}$ -statistic for  $\mu_{r:k}(Q)$  with the degree  $k$  kernel  $h(x_1, \dots, x_k) = x_{r:k}$  is the  $\underline{Q}$ -statistic (Kaigh, 1988 or Kaigh and Driscoll, 1987)

$$M_{r:k;n} = \sum_{1 \leq j \leq n-1} \left[ \binom{j-1}{r-1} \binom{n-j}{k-r} / \binom{n}{k} \right] X_{j:n}.$$

It follows that  $\underline{U}$ -statistic estimators  $h_{k;n}$  of the Fourier coefficients in (1) are then

$$h_{k;n} = (2k+1)^{1/2} (k+1)^{-1} \sum_{0 \leq r \leq k} (-1)^{k-r} \binom{k}{r} (k-r+1) (M_{r+1:k+1;n} - M_{r:k+1;n}).$$



Further calculation with the statistic in (13) produces

$$T_{k,n} = \eta_{k,n} h_{k,n} \quad \text{with } \eta_{k,n} = \left[ \binom{2n-1}{n+k} / \binom{2n-1}{n} \right]^{1/2} \quad (16)$$

to yield

$$ET_{k,n} = \eta_{k,n} \int_{0 \leq u \leq 1} \Pi_k(u) (1-u) q(u) du.$$

Because  $\eta_{k,n}$  has limit one as  $n \rightarrow \infty$ , standard results on  $\underline{U}$ -statistics in Hoeffding (1948) or Serfling (1980) provide

**Theorem 2.** Suppose  $X_1, \dots, X_n$  are iid random variables on  $[0, \infty)$  with qdf  $q$  and finite variance. Let

$$T_{k,n} = n^{-1/2} \sum_{1 \leq j \leq n} \pi_{k,n-1}(j) (n-j+1) (X_{j:n} - X_{j-1:n})$$

$$\mu_k(q) = \int_{0 \leq u \leq 1} \Pi_k(u) (1-u) q(u) du$$

$$\sigma_k^2(q) = \iint_{0 \leq u, v \leq 1} (u \wedge v - uv) q(u) q(v) (d/du)[(1-u) \Pi_k(u)] (d/dv)[(1-v) \Pi_k(v)] dudv.$$

For each fixed  $k$  as  $n \rightarrow \infty$ ,  $n^{1/2}(T_{k,n} - \mu_k(q)) \Rightarrow N(0, \sigma_k^2(q))$ .

Elementary but somewhat tedious calculation shows that the asymptotic variance integral expression in Theorem 2 has the appropriate value  $\beta^2$  for exponential distributions.

It follows also that  $T_{k,n} / \bar{X}$  converges with probability one to the Fourier coefficient of the scaled total time on test functional from (2). Essentially a standardized ratio of  $\underline{U}$ -statistics, the scale-free component  $Z_{k,n} = -(n+1)^{1/2} \eta_{k,n} (h_{k,n} / \bar{X})$  from (12) then has asymptotic normal distribution under alternatives (see Hoeffding 1948, Theorem 7.5). Although computable, the three-term asymptotic variance expression which includes the Theorem 2 expression  $\sigma_k^2(q)$  is quite complicated. Because convergence of  $\underline{U}$ -statistics ratios to their limiting normal distributions is typically slow (see Schechtman and Yitzhaki, 1987), the asymptotic alternative distribution of the scale-free TTT components is not explicitly stated here. To assess performance of the statistics (12) for a specific alternative, however, an adequate qualitative approximation is

$$EZ_{k,n}^2 \approx 1 + n \left[ \int_{0 \leq u \leq 1} \Pi_k(u) (1-u)q(u) du / \mu \right]^2.$$

Observe that this approximation is exact under the null hypothesis and that the Fourier coefficient of the total time on test derivative provides the dominant term for alternative distributions.

### 3. MONTE CARLO POWER COMPARISONS

To investigate efficiency properties and asymptotic distribution theory for individual components and the new omnibus statistics  $QA_n^2$  and  $FQA_n^2$ , a Monte Carlo power study with sample sizes  $n=20, 40$  was performed with 10,000 simulated samples from the standard exponential distribution ( $\beta=1$ ) and several alternative distributions. For empirical power comparisons with  $QA_n^2$  and  $FQA_n^2$ , the Gini statistic  $G_n$ , the squared coefficient of variation  $CV_n^2$ , and the EDF Anderson-Darling statistic  $FA_n^2$  were included as well.

Empirical rejection proportions for  $n=20$  only are presented in Table 1. Nominal significance levels employed for the fourteen alternatives in Table 1(a) and for the ten alternatives in Table 1(b) were 0.10 and 0.05, respectively. Interpolation with critical values for the Greenwood statistic in D'Agostino and Stephens (1986, Table 8.3) provided percentage points for  $CV_n^2$ , whereas asymptotic percentage points were employed for the statistics  $G_n$ ,  $FA_n^2$ ,  $QA_n^2$ , and  $FQA_n^2$ . Results not presented here are consistent with Table 1 to indicate that asymptotic distribution theory of individual components provides adequate approximations for small samples.

Probability density functions for all alternative distributions other than the arcsin [cdf  $F(x)=1/2+(1/\pi)\sin^{-1}(2x-1)$ ,  $0 \leq x \leq 1$ ] appear in Gail and Gastwirth (1978b, Table 3) or in Angus (1982). Except for the arcsin, all alternatives have been used in previous power studies by either Gail and Gastwirth(1978a,b), Lin and Mudholkar (1980), Angus (1982), or Rayner and Best (1986, 1989). For convenient cross-referencing, the alternative-significance level configurations in Table 1 facilitate comparisons of rejection proportions here with those for other statistics reported in previous studies.

Overall,  $FA_n^2$ ,  $QA_n^2$ , and  $FQA_n^2$  emerge as the best omnibus test statistics. Results in Table 1

Table 1. Empirical Power Comparisons for  $n = 20$ 

(a) $\alpha = 0.10$					
Distribution	$G_{20}$	$CV_{20}^2$	$QA_{20}^2$	$FA_{20}^2$	$FQA_{20}^2$
$\chi_1^2$	.6487	.3716	.6421	.7482	.7107
$\chi_3^2$	.2910	.1213	.2865	.2740	.2844
$\chi_4^2$	.6148	.2541	.6208	.5857	.6123
$\chi_8^2$	.9957	.8277	.9973	.9948	.9965
Log normal (0.6)	.8881	.7124	.9646	.9177	.9488
Log normal (0.8)	.3574	.2511	.5065	.3999	.4620
Log normal (1.0)	.1852	.1665	.2737	.2173	.2500
Log normal (1.2)	.3791	.2674	.4112	.3809	.3979
Weibull (0.5)	.9434	.7336	.9417	.9671	.9589
Weibull (2.0)	.9787	.6115	.9730	.9695	.9720
Beta (1, 2)	.4046	.1823	.3660	.4198	.3936
$0.5 (\chi_{0.5}^2 + \chi_4^2)$	.5850	.5106	.6902	.9081	.8601
$0.5 (\chi_1^2 + \chi_3^2)$	.2347	.2702	.2786	.4726	.3945
Arcsin	.3769	.5413	.6346	.8659	.7942
Null	.1028	.0981	.0960	.0997	.0979
(b) $\alpha = 0.05$					
Distribution	$G_{20}$	$CV_{20}^2$	$QA_{20}^2$	$FA_{20}^2$	$FQA_{20}^2$
Weibull (0.8)	.2398	.1167	.2320	.2599	.2483
Weibull (1.5)	.4928	.1497	.4863	.4604	.4777
Uniform (0, 2)	.7136	.2674	.6937	.8005	.7583
Pareto (3)	.7954	1.0000	1.0000	.9918	.9992
Shifted Pareto (3)	.4704	.2973	.4797	.4651	.4760
Shifted Exponential (0.2)	.2264	.1451	.3451	.2167	.2846
Gamma (2)	.4672	.1551	.4990	.4443	.4771
$0.5 (\chi_{0.5}^2 + \chi_4^2)$	.4732	.3586	.5521	.8590	.7840
$0.5 (\chi_1^2 + \chi_3^2)$	.1510	.1650	.1677	.3493	.2688
Arcsin	.2735	.3949	.4432	.7647	.6568
Null	.0498	.0500	.0531	.0531	.0522

show that  $CV_n^2$  performs poorly against all alternatives and that the Gini statistic almost always produces fewer rejections than both  $QA_n^2$  and  $A_n^2$  and their hybrid average. In particular,  $G_n$  provides little protection against lognormal and “bathtub” failure rate (BFR) alternatives. The BFR alternatives, which include the arcsin and chisquare mixtures, have hazard rates which initially decrease and then increase. Comparing  $QA_n^2$  and  $A_n^2$  directly, these test statistics produce similar rejection proportions, with  $QA_n^2$  slightly more powerful for a slim majority of the alternatives investigated. Differences in performance are most pronounced for lognormal and BFR alternatives. For lognormal alternatives,  $QA_n^2$  is clearly the more powerful ; although exhibiting considerably more power than the Gini statistic,  $QA_n^2$  offers less protection than  $A_n^2$  against BFR alternatives.

Comparing results with previous power studies,  $QA_n^2$  and  $FQA_n^2$  are certainly competitive with other test statistics proposed in the literature (Angus, 1982; Lin and Mudholkar, 1980; Rayner and Best 1986, 1989). Supported by these Monte Carlo results, we recommend the statistic  $FQA_n^2$  as an omnibus scale-free for exponentiality. Despite never exhibiting power to exceed that of both  $FA_n^2$  and  $QA_n^2$ , hybrid rejection percentages were always closer to the maximum value.

All simulations employed an IBM 4381 (32 bit word). For distributions without closed form quantile functions, IMSL routines were used; all other distributions were sampled by applying the transformation  $\log[1/(1-U)]$  to each of  $n+1$  uniform variates obtained from the random number generator (the  $n$ th uniform is defined by  $U_n = I_n/P$  with  $P=2^{31}-1$  and  $I_n$  generated by the multiplicative congruential generator  $I_n = I_{n-1} * 16807 \bmod P$ ). The logarithms produce a sample of  $n+1$  iid standard exponential random variables, which is then employed to provide  $n$  uniform order statistics without data sorting (Shorack and Wellner, 1986, p. 335). Alternative distribution samples were obtained from uniform order statistics by qf transformation.

#### 4. ASSESSING EXPONENTIALITY: AN EXAMPLE

Illustrating application of the significance tests to real data, we integrate discussion of general TTT descriptive techniques as well. The following twenty bearing operational lifetimes (in hours)

were presented in Angus (1982) and further analyzed by Rayner and Best (1986, 1989):

2398, 2812, 3113, 3212, 3253, 5236, 6215, 6278, 7725, 8604, 9003, 9350, 9460, 11584, 11825, 12628, 12888, 13431, 14266, 17809.

#### 4.1 Components and Significance Tests

Numerical calculation of the first four TTT spacings components is illustrated in Table 2. This computational display of individual inner product terms can identify the source of a large component value. Resulting inner products are then normalized to produce component numerical values. Exploiting a three-term recursive relation (see Kaigh, 1992a, Table 1), generation of Hahn polynomial vectors with integer entries and squared norms  $[1/(2k+1)](n+k)!/(n-k-1)!$  for machine calculation with (12) is quite simple. Because manual input is required for only the initial constant and linear entries, the necessary calculations are easily performed with a simple desktop computer spreadsheet program. Only the first four components are treated in Table 2, but the computational scheme permits calculation of higher order components as well.

Computed values for all nineteen TTT spacings components are

2.97, -1.24, 1.41, -2.32, 2.27, -2.12, .68, .17, .13, -.02, -.04, -1.19, 1.28, .49, -1.24, .57, -.90, -1.09, .10.

As typical, low frequency components capture most of the spacings variation. For these data the first six components account for about 78% of the total sum of component squares value 35.1. To assess statistical significance, individual TTT components are compared to standard normal percentage points. Accordingly, the first component, which is equivalent to the Gini statistic, is highly significant. Aggregating the contributions of individual TTT components, the omnibus statistics values  $CV_n^2 = 1.759$  ( $.025 < p\text{-value} < .05$ ) and  $QA_n^2 = 5.424$  ( $p\text{-value} = .002$ ) also present strong evidence against the null exponentiality assumption. Values for the other quadratic statistics are  $FA_n^2 = 4.912$  ( $p\text{-value} = .003$ ) and  $FQA_n^2 = 5.168$  ( $p\text{-value} = .002$ ). Now as well as later, asymptotic significance probabilities for  $CV_n^2$  were obtained from D'Agostino and Stephens (1986, Table 8.3), whereas those for  $FA_n^2$ ,  $QA_n^2$ , and  $FQA_n^2$  were computed numerically using the algorithm in Martynov (1975).

#### 4.2 TTT Plot and Further Analysis

The upper portion of Figure 1 shows the TTT plot of discrete points  $(j/n, S_{j:n})$ ,  $j=0,1,\dots,n$ ,

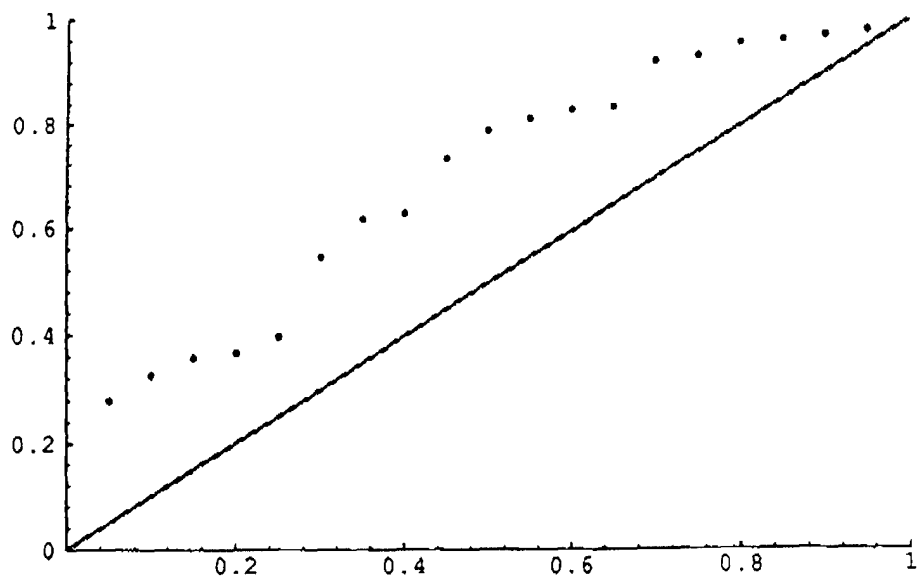
Table 2. Bearing Operational Lives Total Time on Test Data Analysis

TTT Spacings	x	Hahn Polynomial Vectors					Inner Product Terms				
		P <sub>0</sub>	P <sub>1</sub>	P <sub>2</sub>	P <sub>3</sub>	P <sub>4</sub>	0	1	2	3	4
.28	1	1	-19	342	-5814	93024	.28	-5.32	95.77	-1628.07	26049.14
.05	2	1	-17	234	-2142	-4896	.05	-.78	10.74	-98.33	-224.75
.03	3	1	-15	138	510	-53856	.03	-.47	4.37	16.15	-1704.92
.01	4	1	-13	54	2262	-67296	.01	-.13	.53	22.23	-661.32
.03	5	1	-11	-18	3234	-56976	.03	-.32	-.52	93.96	-1655.38
.15	6	1	-9	-78	3546	-32976	.15	-1.35	-11.70	531.99	-4947.26
.07	7	1	-7	-126	3318	-3696	.07	-.51	-9.16	241.12	-268.58
.01	8	1	-5	-162	2670	24144	.01	-.06	-1.88	31.01	280.38
.10	9	1	-3	-186	1722	45504	.10	-.30	-18.86	174.58	4613.38
.06	10	1	-1	-198	594	57024	.06	-.06	-11.18	33.53	3219.23
.02	11	1	1	-198	-594	57024	.02	.02	-4.61	-13.84	1328.49
.02	12	1	3	-186	-1722	45504	.02	.05	-3.39	-31.40	829.72
.01	13	1	5	-162	-2670	24144	.01	.03	-.83	-13.72	124.05
.09	14	1	7	-126	-3318	-3696	.09	.61	-10.94	-288.04	-320.85
.01	15	1	9	-78	-3546	-32976	.01	.08	-.66	-29.94	-278.42
.02	16	1	11	-18	-3234	-56976	.02	.26	-.42	-75.81	-1335.63
.01	17	1	13	54	-2262	-67296	.01	.08	.33	-13.73	-408.62
.01	18	1	15	138	-510	-53856	.01	.14	1.31	-4.85	-512.28
.01	19	1	17	234	2142	-4896	.01	.17	2.28	20.88	-47.74
.02	20	1	19	342	5814	93024	.02	.39	7.07	120.27	1924.39
Inner Products							1	-7.48	48.25	-912.00	26003.04
Components								2.97	-1.24	1.41	-2.32

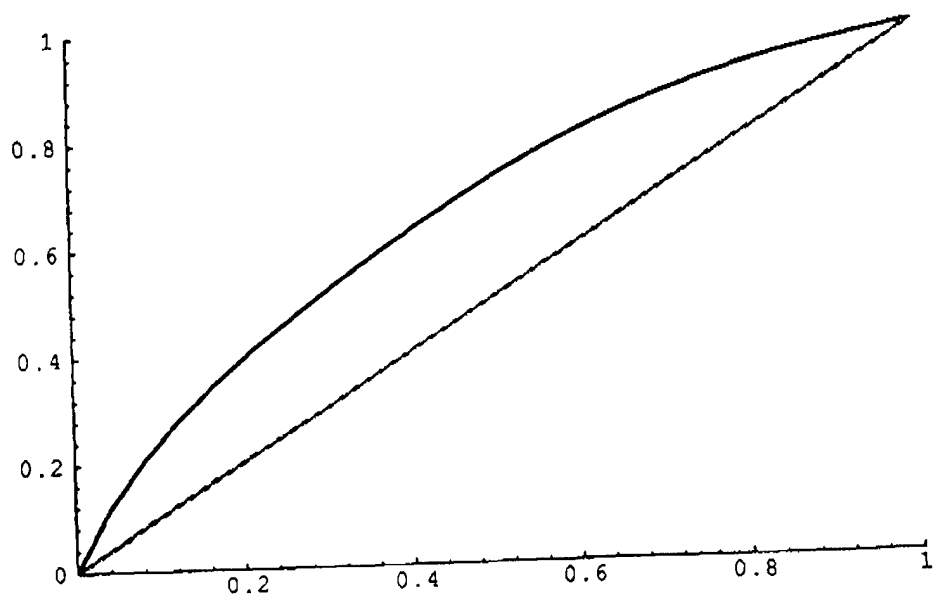
NOTE: Components are calculated from the corresponding spacings inner products using the formula  

$$Z_{k,n} = - [n(n+1)]^{1/2} (2k+1)^{1/2} [(n-k-1)!/(n+k)!]^{1/2} (\mathbf{p}_k^T \mathbf{D}).$$

Figure 1. Scaled Total Time on Test Plots for Bearing Lifetime Data



TTT Plot



Smooth TTT Plot

formed for the original twenty bearing lifetimes. For visual comparison corresponding null expected values appear along the 45° line. Although noisy, this plot strongly suggests a shifted exponential distribution. Returning briefly to Table 2, observe the effect of the first TTT spacing with value .28 on the magnitudes of the corresponding inner product terms in the first row. This single large spacing is mainly responsible for the magnitudes of the low-frequency components.

In general, simply subtracting the minimum from each of the original data values produces an exponential sample of  $n-1$  (D'Agostino and Stephens, 1986, p. 425). The TTT procedures developed in Sections 1 and 2 are then easily adapted to assess the more general model that  $F$  is exponential with shift parameter  $\gamma$  and mean  $\beta$  both unspecified.

Subtracting the minimum lifetime 2398 from each original observation and recalculating using the shifted lifetime data yields eighteen new component values  
1.63, 1.00, -1.17, -.05, .58, -1.46, .52, .48, -.31, .73, -1.62, .12, 1.92, -1.47, .34, -.57, -1.69, -.11,  
none of which is significant individually. Calculated values for the aggregate statistics with  $n=19$  are now  $CV_n^2 = 1.138$  ( $p\text{-value} > 0.10$ ),  $FA_n^2 = 2.012$  ( $p\text{-value} = .090$ ),  $QA_n^2 = 1.748$  ( $p\text{-value} = .119$ ), and  $FQA_n^2 = 1.880$  ( $p\text{-value} = .106$ ). Although this TTT analysis also casts doubt on exponentiality, the results are not statistically significant.

### 4.3 Continuous TTT Plots

The TTT plot is generally useful for many applications in reliability including model identification and age replacement theory (see Bergman and Kelfsjö, 1984). Similar treatment applies to the shifted data, but we address now only plots of the original twenty bearing lifetimes.

Although the upper portion of Figure 1 is consistent with the preceding EQF analyses, noise obscures the effect. For presentation purposes we recommend the continuous plot displayed in the lower portion of Figure 1. This graph was obtained from the Bernstein-like polynomial

$$S^*(t) = \sum_{0 \leq r \leq n} S_{r:n}^* \binom{n}{r} t^r (1-t)^{n-r}, \quad 0 \leq t \leq 1,$$

with

$$S_{r:n}^* = \sum_{0 \leq j \leq n} \left[ \binom{r+j-1}{r-1} \binom{2n-j-r-1}{n-r-1} \binom{2n-1}{n} \right] S_{j:n}, \quad 1 \leq r \leq n-1, \quad S_{0:n}^* = 0 \text{ and } S_{n:n}^* = 1.$$

Generalized order statistics  $S_{r:n}^*$  from Kaigh and Cheng (1991) provide the discrete input to the



Bernstein operator which then produces a nondecreasing function with the appropriate endpoint values zero and one. The continuous function  $S^*(t)$  would also inherit certain other discrete TTT properties including symmetry or concavity. In addition, Legendre and Hahn polynomial relations in Kaigh (1992a) establish the component representation

$$\int_{0 \leq t \leq 1} [S^*(t)-t]^2 [t(1-t)]^{-1} dt = (n+1)^{-1} \sum_{1 \leq k \leq n-1} [1/k(k+1)] \eta_{k,n}^6 Z_{k,n}^2$$

with  $\eta_{k,n}$  as in (16). Similar to the decomposition (15) but with more pronounced emphasis on low-order components, this representation demonstrates that the continuous version of the TTT also can be used to quantitatively assess exponentiality.

Although not developed here, application of the Bernstein operator to the original TTT values (5) also yields a continuous, but somewhat more irregular, TTT plot with weighted squared norm obtained by simply substituting  $\eta_{k,n}^2$  for  $\eta_{k,n}^6$  in the decomposition above. It follows that high-frequency contributions to the recommended  $S^*(t)$  are damped considerably more.

Inherent characteristics of sample TTT values are preserved by the isotonic linear transformation with matrix representation specified in the defining expression for  $S_{r:n}^*$ , but with reduced variability. The Lorenz partial order majorization results in Kaigh and Sorto (1992) justify mathematically our assertion that the continuous TTT plot is smoother than the original. This fact is clearly illustrated in Figure 1.

#### 4.4 TTT Numerical Summaries

To further enhance data analysis and description it is useful to have a concise set of numbers which summarize the TTT. In this light, we suggest a four-number summary of  $\underline{Q}$ -statistics employing TTT values (5).

All TTT spacings components of order  $k$  or less can be obtained from the collection of  $\underline{Q}$ -statistics defined for the TTT by

$$S_{r;k;n} = \sum_{1 \leq j \leq n-1} \binom{j-1}{r-1} \binom{n-j-1}{k-r} \binom{n-1}{k} S_{j:n}, \quad 1 \leq r \leq k.$$

Related to the  $\underline{L}$ -moments in Hosking (1990), these statistics treat the TTT values and not the original data as previously in Section 2.2. Calculations in Kaigh (1988) show that  $S_{r;k;n}$  is a  $\underline{U}$ -

statistic with mean  $r/(k+1)$  for uniform data. The TTT spacings components (12) are standardized  $\underline{U}$ -statistics (in the transformed data) with representation as linear combinations of  $\underline{Q}$ -statistics

$$Z_{k,n} = (2k+1)^{1/2} (n-1)^{1/2} \left[ \binom{2n-1}{n+k} / \binom{2n-1}{n+1} \right]^{1/2} \sum_{1 \leq r \leq k} (-1)^{k-r} \binom{k+1}{r} [S_{r:k;n} - r/(k+1)] .$$

The location, scale, skewness, and kurtosis TTT components ( $Z_{1,n}, Z_{2,n}, Z_{3,n}, Z_{4,n}$ ) are then algebraically equivalent to the  $\underline{Q}$ -statistic four-number summary  $[S_{1:4;n}, S_{2:4;n}, S_{3:4;n}, S_{4:4;n}]$  corresponding to the null mean vector  $[.2, .4, .6, .8]$ . Thus, simple comparison of the  $\underline{Q}$ -statistics to the null mean vector can augment informally results from the significance tests and TTT plots. In addition, the  $\underline{Q}$ -statistic summary provides a convenient method for comparing TTT plots from samples of disparate sizes. To bypass the defining expression given above, a simple recursive computational formula (Kaigh, 1988 or Kaigh and Driscoll, 1987) permits rapid machine calculation of all  $n(n-1)/2$   $\underline{Q}$ -statistics for the TTT values (5).

Returning to the bearing lifetimes a final time,  $\underline{Q}$ -statistic summaries for the original data and the shifted data are  $[.42, .63, .81, .93]$  and  $[.24, .53, .76, .91]$ , respectively. In agreement with the formal significance tests and the TTT plots, deviations of the shifted data summary from the null values  $[.2, .4, .6, .8]$  are considerably smaller.

## 5. CONCLUDING REMARKS

Application of Fourier type methods to the TTT yields omnibus as well as directional scale-free tests for exponentiality. The asymptotically normal TTT spacings components are essentially point estimators of Legendre polynomial Fourier coefficients of the total time on test transform derivative. Related to the  $\underline{L}$ -moments in Hosking (1990), the first four components quantify location, scale, skewness, and kurtosis departures from exponentiality. Utilizing readily available EDF Anderson-Darling asymptotic critical points, the TTT aggregate statistics exhibit good power against various exponential alternatives. The recommended hybrid average of both Anderson-Darling type statistics effectively integrates EQF information with that from the EDF. A polynomial TTT plot provides a useful display for continuous data. A four-number summary of  $\underline{Q}$ -statistics is proposed for description. Spacings components here with the TTT complement related goodness-of-fit measures for the normal distribution recommended by Hosking (1992).

## REFERENCES

- Angus, J. E. (1982), "Goodness-of-Fit Tests for Exponentiality Based on a Loss-of-Memory Type Functional Equation," Journal of Statistical Planning and Inference, 6, 241-251.
- Barlow, R. E. (1979), "Geometry of the Total Time on Test Transform," Naval Research Logistics Quarterly, 26, 393-402.
- Bergman, B. and Klefsjo, B. (1982), "The Total Time on Test Concept and Its Use in Reliability Theory", Operations Research, 32, 596-606.
- Chandra, M. and Singpurwalla, N. D.(1981), "Relationships Between Some Notions Which Are Common to Reliability Theory and Economics," Mathematics of Operations Research, 6, 113-121.
- Currie, I. D. and Stephens, M. A. (1986), "Relations Between Statistics for Testing Exponentiality and Uniformity," The Canadian Journal of Statistics, 14, 177-180.
- D'Agostino and Stephens, M. A. (eds.) (1986), Goodness-of-Fit Techniques, Marcel Dekker, New York.
- Durbin, J. and Knott, M. (1972), "Components of Cramer-von Mises Statistics I," Journal of the Royal Statistical Society, Ser. B, 34, 290-307.
- Gail, M. H. and Gastwirth, J. L.(1978a), "A Scale-Free Goodness-of-Fit Test for the Exponential Distribution Based on the Gini Statistic," Journal of the Royal Statistical Society, 40, 350-357.
- \_\_\_\_\_ (1978b), "A Scale-Free Goodness-of-Fit Test for the Exponential Distribution Based on the Lorenz Curve," Journal of the American Statistical Association, 73, 787-793.
- Greenwood, M. (1946), "The Statistical Study of Infectious Diseases," Journal of the Royal Statistical Society, Ser. A, 109, 85-110.
- Hoeffding, W. (1948), "A Class of Statistics With Asymptotically Normal Distribution," Annals of Mathematical Statistics, 19, 293-325.
- Hosking, J. R. M. (1990), "L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics," Journal of the Royal Statistical Society, Ser. B, 52, 105-124.

Hosking, J. R. M. (1992), "Moments or  $L$  Moments? An Example Comparing Two Measures of Distributional Shape," The American Statistician, 46, 186-189.

Kaigh, W. D. (1988), "Q-Statistics and Their Applications," Communications in Statistics, Part A-Theory and Methods, 17, 2191-2210.

\_\_\_\_\_ (1992a), "EDF and EQF Orthogonal Component Decompositions," Journal of Nonparametric Statistics. In press.

\_\_\_\_\_ (1992b), "Distribution-Free Two-Sample Tests Based on Rank Spacings." Unpublished manuscript.

Kaigh, W. D. and Cheng, C. (1991), "Subsampling Quantile Estimators and Uniformity Criteria," Communications in Statistics, Part A-Theory and Methods, 20, 539-560.

Kaigh, W. D. and Driscoll, M. F. (1987), "Numerical and Graphical Data Summary Using Q-Statistics," The American Statistician, 41, 25-32.

Kaigh, W. D. and Sorto, M. A. (1992), "Subsampling Quantile Estimator Majorization Inequalities." Unpublished manuscript.

Lin, C. C., and Mudholkar (1980), "A Test of Exponentiality Based on the Bivariate F Distribution," Technometrics, 22, 79-82.

Martynov, G. V. (1975), "Computation of Distribution Functions of Quadratic Forms of Normally Distributed Random Variables," Theory of Probability and Its Applications, 20, 782-793.

Neuman, C. P. and Schonbach, D. I. (1974), "Discrete (Legendre) Orthogonal Polynomials-A Survey," International Journal for Numerical Methods in Engineering, 8, 743-770.

Nikiforov, A. F., Suslov, S. K., and Uvarov, V. B. (1991), Classical Orthogonal Polynomials of A Discrete Variable, Springer-Verlag, Berlin.

Rayner, J. C. W. and Best, D. J. (1986), "Neyman-Type Smooth Tests for Location-Scale Families," Biometrika, 73, 437-446.

\_\_\_\_\_ (1989), Smooth Tests of Goodness of Fit, Oxford University Press, New York.

Schechtman, E. and Yitzhaki, S. (1987), "A Measure of Association Based on Gini's Mean Difference," Communications in Statistics, Part A-Theory and Methods, 16, 207-231.

Serfling, R. J. (1980), Approximation Theorems of Mathematical Statistics, John Wiley & Sons, New York.

Shorack, G. R. and Wellner, J. A. (1986), Empirical Processes With Applications to Statistics, John Wiley & Sons, New York.

DETERMINATION OF THE ECONOMIC ACCEPTABLE QUALITY LEVEL (EAQL)  
J. STEVE CARUSO, INDUSTRIAL ENGINEER  
U.S. ARMY MANAGEMENT ENGINEERING COLLEGE  
ROCK ISLAND, IL 61299-7040

ABSTRACT

This paper presents an empirical formula that can be utilized to establish a mathematical determination of the AQL, Acceptable Quality Level. The AQL is arrived at by determining an Economic Acceptable Quality Level (EAQL). A formula is provided as an estimate of the EAQL on the premise of a worst case scenario, and integrally requires the use of a sampling standard incorporating a family of sampling plans. The formula is developed using economic (cost) considerations and equates the cost of inspection to the cost of correcting deficiencies that may pass the inspection station. The sampling standard utilized is ANSI/ASQC Z1.4, (MIL-STD-105E), due to a unique mathematical relationship that exists where the  $3(AQL) = AOQL$  for the  $Ac=0$  plans. The formula:

$$EAQL = \frac{(n)(CI)}{(3)(N)(CR)} \text{ , where } AQL = EAQL$$

- n = Sample size determined using ANSI/ASQC Z1.4, (MIL-STD-105E) Standard.  
N = Lot size  
CI = Cost of inspecting one unit.  
CR = Cost to correct the deficiency on one unit of product.

INTRODUCTION

Specifying an AQL, Acceptable Quality Level, to be incorporated into a specification, determined on a quantitative basis, has been a long standing problem. Selecting an appropriate AQL in a factory for auditing purposes, determined on a quantitative basis, has similarly been an unresolved problem. The purpose of this paper is to provide a method of determining a suitable AQL in these situations.

To quote J.W. Wiesen (Reference 4), "The AQL can not be set scientifically and hence must be set by bargaining or arbitrarily." In this same article, Wiesen mentions Enell's decision model (Reference 2). Enell's model provides a basis for deciding whether to sort or accept a lot without inspection utilizing the known

production incoming fraction defective ( $p$ ). Enell's model is

$$P_b = \frac{I}{A}$$

Where  $I$  = Cost to Inspect One Unit and  $A$  = Damage Done By One Defective. Lots having an incoming fraction defective quality level less than  $P_b$  should be accepted without sorting, and those having an incoming quality level above  $P_b$  should be sorted. This model is acceptable for decisions involving sorting. The model is restrictive in that it requires prior knowledge of the process average (incoming quality level) in order to arrive at a decision by quality personnel as to whether to sort or accept the material as is. In attempting to use this approach in a specification for material to be delivered in the future, or in an in-plant situation where manufacturing has not commenced, the model is not very useful.

#### DEVELOPING THE MODEL

Two factors contribute to an empirical approach in the selection of an AQL:

1. An established sampling standard incorporating a family of sampling plans.
2. An established sequence of steps to be followed when administering the standard while sampling.

The first requirement is satisfied through the use of ANSI/ASQC Z1.4 (MIL-STD-105E) "Sampling Procedures And Tables For Inspection By Attributes" (Reference 1). It is possible that another published or developed standard could be used. However, this standard was chosen because of its wide dissemination and the unique empirical relationship that exists between AQLs and AOQLs for the  $A_c = 0$

plans. Table V-A of the standard displays a consistent relationship for these plans of  $3(AQL) = AOQL$ . The second requirement is satisfied within the standard as follows (see Figure 1):

Incoming Lot No. 1 (e.g. on rejection). The lot is presented under original inspection. The entire sample quantity ( $n$ ) is inspected. On rejection, the defective items are counted. Identify the defective quantity as  $D$ . Calculate the estimated portion defective

" $p$ " using  $p = \frac{D}{n}$  where  $n$  represents the sample size.

This calculated value of  $p$  is then compared to  $P_b$  in Enell's model to determine whether the remainder of the lot quantity,  $N-n$ , will be sorted or accepted as is. Thus, this remaining quantity,  $N-n$ , if screened, is precipitated by a separate independent decision. Therefore, in rejected lots, the sampled quantity,  $n$ , is considered as the total quantity of material inspected. Incoming Lot No. 2 (e.g. on acceptance). The only quantity of items inspected in lots accepted on original inspection is the sampled quantity ( $n$ ), see Figure 1. Thus, for both rejected and accepted lots,  $n$  represents the total number of units inspected per lot.

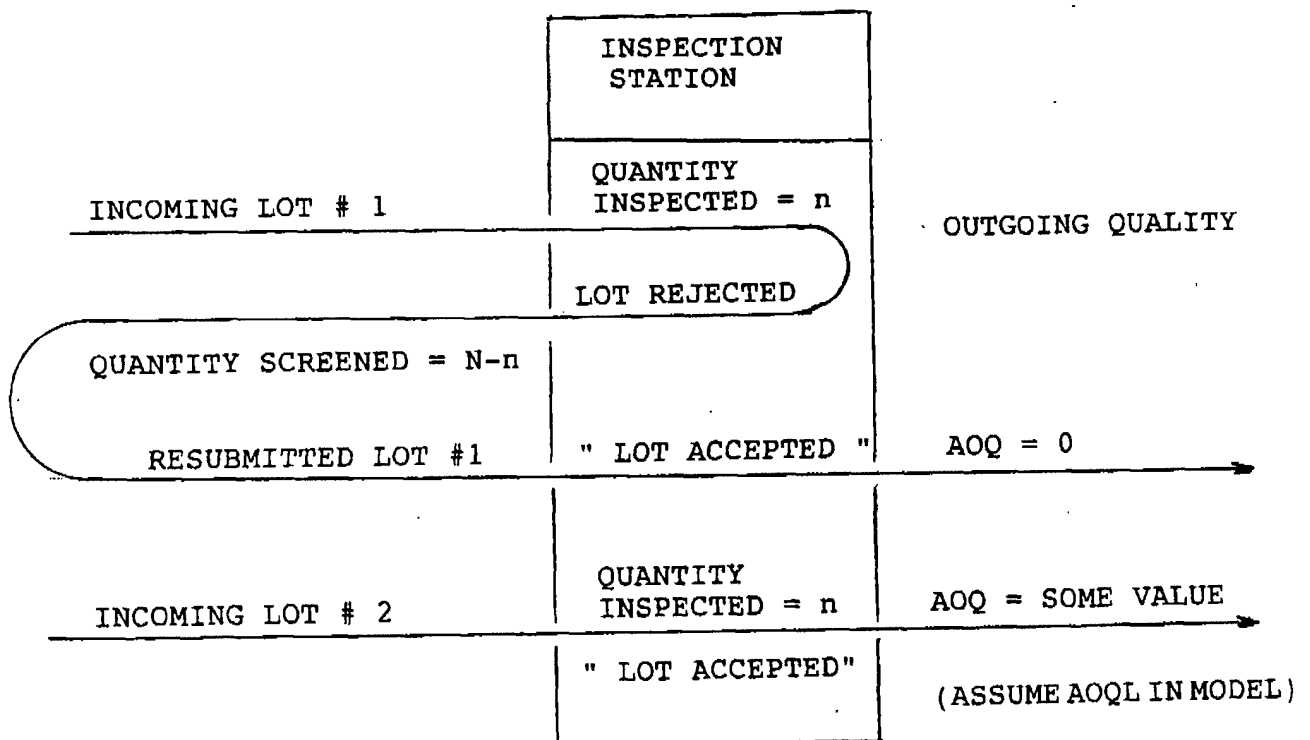


Figure 1. OPERATION OF INSPECTION STATION



The model will require an estimate of the total quantity of defective items which get past the inspection station. The best estimate of this quantity is the Average Outgoing Quality (AOQ). However, every AOQ relates to a specific value of incoming fraction defective. If we elect to establish a known value of incoming fraction defective, we could revert to Enell's model. Then by comparing our value of incoming fraction defective to the

calculated  $P_b$ , we can decide whether to sort the material or accept "as is". There is a way out of the dilemma. For any sampling plan, we assume the worst condition of AOQ which is the Average Outgoing Quality Limit (AOQL) and it represents the average amount of unacceptable material which passes the inspection station per lot regardless of the incoming rate of defective material.

These are the two essential theoretical considerations necessary to develop the model. The number of items to be inspected per lot and the average percent of defective material which gets past the inspection station per lot. The model is further developed under the assumption of perfect inspection.

In order to scientifically establish the AQL, there are two costs that must be considered. These two costs are the total cost of inspection (labor, materials, equipment, setup, etc) and the total cost (or damage) to correct the defects which pass the inspection station. These latter costs are the in-plant failure costs and may include external failure costs when the item is delivered outside the plant. The in-plant costs relate to elements such as scrap, rework, reassembly, etc. The external failure costs incorporate elements such as liability, replacement, service cost, warranty, and goodwill.

The model is developed by calculating the indifference cost. The indifference cost occurs at the point where the total cost of inspection equals the total cost to correct the defects which pass the inspection station. The model is designed to provide a best mathematical approximation of the acceptable incoming fraction defective at the indifference point. The EAQL is equated to this incoming fraction defective. The EAQL will then represent that level of incoming fraction defective that for purposes of sampling can be assumed satisfactory. The equation will also automatically identify those characteristics where 100% inspection must be preformed. As the calculated EAQL approaches or equals 0 no defective material is authorized to pass the inspection station.

Development of the indifference equation:

Total cost of sampling inspection per lot =  $n \times CI$ .

Total cost of correcting deficiencies per lot =  $AOQL \times N \times CR$ .

Where:  $n$  = sample size

$N$  = lot size

$I$  = total cost in dollars to inspect one unit

CR = total cost in dollars to correct the defective item (internal and external considerations)

AOQL = maximum average outgoing quality limit in percent for a sampling plan

The equation:  $n \times CI = AOQL \times N \times CR$

For the  $Ac=0$   $Re=1$  plans in ANSI/ASQC Standard Z1.4 there is a unique relationship between the AQL and the AOQL for the normal plans. That relationship is that  $3 \times AQL = AOQL$ .

The equation is altered to:  $n \times CI = 3 \times AQL \times CR$

Solve for AQL

$$EAQL = \frac{(n) (CI)}{(3) (N) (CR)}$$

AQL in percent

$$EAQL\% = \frac{(n) (CI)}{(3) (N) (CR)} \times 100\%$$

In order to determine the AQL using this equation one must first establish the sample size (n). The steps followed in selecting the sample size are those normally followed when using ANSI/ASQC Z1.4 (MIL-STD-105E). First determine the code letter. Enter the "Sample Size Code Letters" table with the lot size (N) using general level II TO identify the code letter. With the code letter enter the "Single Sampling Plans For Normal Inspection" table to select the associated sample size. The next step is to calculate the EAQL using the equation and the sample size found in the standard. The final step is to select a listed AQL from the standard for a plan having an  $Ac=0$ ,  $Re=1$  values which is  $\leq$  to the calculated EAQL. To meet this requirement it may be necessary move to a different, larger sample size, than was used in the formula to calculate the EAQL.

#### PROCEDURE

The following information must be specified in order to determine the AQL:

Lot Size (N)

Cost To Inspect One Unit (CI)

Cost TO Correct The Defective Item (CR)

The cost of inspection (CI) should incorporate all identifiable charges associated with inspecting one item. It should include all direct and indirect labor and material costs.

The cost to correct the defective item (CR) will incorporate either internal failure costs or external failure costs and in some cases cost elements from both categories will be incurred.

Internal failure costs relate to those charges incurred within the plant to correct the defective item such as rework, repair, reinspection, replacement, delays in production, and additional material handling. External failure costs incorporate charges for injury, warranty replacement, transportation, liability, and goodwill. The cost elements identified in the internal and external failure categories are not easy to estimate but need to be estimated when sampling is to be incorporated into the quality program. The following examples provide the sequence of steps that must be followed in determining the EAQL.

#### EAQL Determination Example Number 1

Basic Information Required For EAQL Determination:

Lot Size N = 1000

CI = \$.10 per unit

CR = \$5.00 per unit

#### Step Number 1

Enter the standard with the lot size N = 1000 and use General Inspection Level II to determine the Code Letter in the "Sample Size Code Letters" table. The code letter is "J". From the "Single Sampling Plans For Normal Inspection" table locate the sample size associated with code letter "J". The sample size is (n)=80.

#### Step number 2

Calculate the EAQL using the formula:

$$EAQL = \frac{(n)(CI)}{(3)(N)(CR)} \times 100\%$$

$$EAQL\% = \frac{(80)(\$ .10)}{(3)(1000)(\$5.00)} \times 100\% = .05\%$$

#### Step Number 3

Establish the EAQL associated with an available sampling plan in ANSI/ASQC Z1.4 (MIL-STD-105E). Reenter the "Single Sampling Plans For Normal Inspection" table in the standard. Enter the table with the established code letter and associated sample size (J, n=80) and read to the Ac=0 Re=1 plan. The AQL identifying this plan must have a value .05% or less. In this case the associated AQL is .15% which is substantially larger than our calculated value of .05%. Therefore, move down the standard AQL values to find an Ac=0 plan with an associated AQL value of .05% or less. The next smallest available AQL value found is .04% for the AC=0 plan, Code Letter M, with an associated sample size of n=315. Thus, the EAQL to be specified is .04%. The sample size of 315 is much larger than 80 initially used in the formula for calculating the EAQL. However, even though the sample size is much larger than economically required it still allows for sampling rather than 100 % inspection.

In the event that the AQL is to be specified into a requirements document; there are two ways it can be done. The following are two methods of specifying the AQL (EAQL):

1. Inspection lot size (N) = 1000  
Normal single  
Level II  
AQL = .04%
2. Inspection lot size (N) = 1000  
Normal single  
Sample size (n) = 315  
Ac = 0, Re = 1

The second method avoids any reference to the AQL.

A comment regarding how closely the cost elements balance in the equation using the sample size and AQL values determined from the standard in satisfying the EAQL formula calculation. The total cost of inspection, (n)(CI) or (315)(\$10) = \$31.50, does not equal the total cost to correct the defective items, (3)(AQL)(N)(CR) or (3)(.0004)(1000)(\$5.00) = \$6.00. This discrepancy occurred because the calculated EAQL is not an available AQL in the standard. In part, the discrepancy also occurred due to the fact that the standard sampling plan sample size was inadequate for the calculated EAQL. The example is a worse case scenario where both initial determinations are modified in satisfying the procedure using the standard.

#### EAQL Determination Example Number 2

Basic Information Required For EAQL Determination:

Lot Size = 2500  
CI = \$15.00  
CR = \$250.00

#### Step Number 1

Enter the standard with the lot size N = 2500 and Inspection Level II and find the Code Letter and associated sample. The code letter is K. The sample size determined from the "single Sampling Plans For Normal Inspection" table is n = 125.

Step Number 2 Calculate the EAQL

$$\text{EAQL}\% = \frac{(n)(CI)}{(3)(N)(CR)} \times 100\% = \frac{(125)(\$15.00)}{(3)(2500)(\$250.00)} \times 100\% = .10\%$$

This EAQL happens to be equivalent to an available AQL in the standard. The sample size required for this lot size and associated calculated EAQL conform to an available Ac=0 plan in the standard. In this case the total cost of inspection versus the total cost of correcting deficiencies should very nearly balance. Total cost of inspection, (n)(CI) or (125)(\$15.00) = \$1875.00 is nearly equal to the total cost of correcting deficiencies, (AOQL)(N)(CR) or (.0029)(2500)(\$250.00) = \$1812.50. The AOQL rather than 3(AQL) was

used in the previous expression since the AOQL was available in the standard and  $3(AQL) = AOQL$ . These two total cost elements will not always balance. These costs will equate when the calculated EAQL equals a standard AQL for a  $Ac=0$  plan in the standard as was the case here.

#### CONCLUDING COMMENTS

The technique outlined in this paper provides a method of arriving at an AQL(EAQL) either to be incorporated into contractual documentation or for process evaluation on the plant floor where no previous knowledge of the process defective rate exists.

The approach as outlined has at least three benefits. The methodology generally produces conservative AQLs. the AQLs are conservative because the EAQL is calculated using the AOQL or the maximum rate of outgoing defective for the sampling plan. Operating at any incoming defective rate other than that which produces the AOQL generates an AOQ below the AOQL. Also when the calculated EAQL and selected sample size do not intersect at an available sampling plan the recommendation is to select a tighter AQL which generally leads to greater inspection than required by the economic model. The method also requires that costs particularly the cost consequences of a bad item in the system be considered. In general, these costs are ignored because they are difficult to estimate but are essential for effective decision making. In some cases the model may indicate that 100% inspection is required. The consequences of allowing even one defective to get by is uneconomical. Thirdly, it provides the user with a procedural quantitative approximation in establishing the AQL when sampling is to be employed.

The general assumption associated with the determination and use of the AQL and its associated sample size is that the lot will be assembled and the sample selected, inspected, and the findings matched to the acceptance criteria for approval or rejection. Some quality practitioners may have trouble accepting the thrust of the paper because they associate the AQL with this historical method of acceptance sampling versus the use of control charts. There are, however, situations which arise where sampling may still be utilized. Auditing, for instance, whether for purchased goods or internal quality verification will still normally require the use of sampling. Processes that, for whatever reason, are not currently being tracked with control charts could also utilize sampling if deemed appropriate.

However, this is just one of the uses of this methodology. The methodology can also be applied even when variable control charts are used. Material produced within a specific time period, such as a day, can be considered the lot and identified as the "quantity produced in the production interval". Follow the procedure to

determine the sample size and AQL. The sample size can then be subdivided into the number of subgroups which exhaust the total sample size for the production interval. Every time a defect is found, while using the control chart, "the quality system is challenged". The challenge requires screening of material produced and to be produced within that production interval. The standard variable control chart procedures can be followed including using the standard tests. However, as indicated, an item falling out of specification is a rejection and requires lot screening.

The use of this AQL determination and application in attribute control charts is not really effective. Attribute control charts normally require a subgroup sample size of at least 50. Therefore, the sample size determined under the procedure will be the subgroup sample size. The central line can be equated to the EAQL and the associated upper and lower control limits calculated for the control chart. Though this upper control limit can be established it is meaningless. The  $Ac$  (acceptance number) = 0 requires a percent defective "p" of 0% for the subgroup. Therefore, problem detection is signaled by a defect not position or run conditions on the chart. For informational purposes the chart can be constructed and lot fraction can be tracked on the chart. The fraction defective can be calculated either from the information generated from the subgroup sample size or after the material is screened which would provide a more accurate estimate of fraction defective. The chart thus indicates quality status.

DETRMINATION OF THE ECONOMIC ACCEPTABLE QUALITY LEVEL  
J. STEVE CARUSO, INDUSTRIAL ENGINEER  
U.S. ARMY MANAGEMENT COLLEGE  
ROCK ISLAND, IL 61299-7040  
TELEPHONE (309) 782-0509

Terms:

- n = The sample size found using the appropriate code letter based on the inspection lot size using General Inspection Level II and Normal Inspection in ANSI/ASQC Z1.4, (MIL-STD-105E) Standard.
- N = Inspection Lot Size (Must be specified in order that the sample size be selected).
- AQL = The acceptable quality level expressed which, for the purposes of sampling inspection, can be considered the limit of satisfactory process average.
- EAQL= Economic Acceptable quality level equated to the AQL and determined using cost considerations.
- CI = Total cost in dollars to inspect one unit (materials, labor, Instrumentation, etc.)
- CR = Total cost in dollars to repair or replace a defective unit. For work in process, total internal cost; for field failures, total external plus any internal cost.
- Ac = Acceptance Number, the maximum number of defects or defectives found in the sample which if not exceeded allows acceptance of the lot.
- Re = Rejection Number, lot rejection is recommended when this number of defects or defectives is found in the sample.

DETERMINATION OF THE ECONOMIC ACCEPTABLE QUALITY LEVEL (EAQL)  
J. STEVE CARUSO, INDUSTRIAL ENGINEER  
U.S. ARMY MANAGEMENT ENGINEERING COLLEGE  
ROCK ISLAND IL, 61299-7040  
TELEPHONE (309) 782-0509

REFERENCES

1. ANSI/ASQC Z1.4-1981 (1981) "Sampling Procedures And Tables For Inspection By Attributes." American National Standards Institute/American Society For Quality Control, 1981; military equivalent MIL-STD-105E.
2. Enell, J.W. (1954) "Which Sampling Plan Should I Choose," Industrial Quality Control, May 1954, pp. 96-100.
3. Peterson, C. (1970) "Selecting A Product Quality Level," Industrial Engineering, August 1970, pp. 23-26
4. Wisen, J.W. (1951) "Sampling By Attributes," Juran's Quality Control Handbook, pp. 24-1 thru 24-44.
5. Masser, W.J. (1957) "The Quality Manager And Quality Costs," Industrial Quality Control, October 1957.
6. Hurayt, G. (1989) "Estimating AOQL For Zero Nonconformance," Quality, October 1989, pp. 49.



DETERMINATION OF THE ECONOMIC ACCEPTABLE QUALITY LEVEL (EAQL)

J. STEVE CARUSO, INDUSTRIAL ENGINEER  
U.S. ARMY MANAGEMENT ENGINEERING COLLEGE  
ROCK ISLAND, IL 61299-7040  
TELEPHONE (309) 782-0509

KEY WORDS

Acceptable Quality Level

Economic Quality Level

Economic Acceptable Quality Level

Sampling Inspection

Determination of Sampling Plan

Acceptance Inspection

# Some Problems of Estimation and Testing in Multivariate Statistical Process Control

*Martin Lawera*  
*James R. Thompson*  
Department of Statistics  
Rice University  
Houston, TX 77251-1892

January 25, 1993

## Abstract

In the search for Pareto glitches, almost all testing is done one variable at a time. In many situations, however, the time-indexed observables are vectors. A procedure for obtaining trimmed multivariate estimates of location is presented. We propose various parametric testing procedures to exploit some of the special structures generally present in a quality control setting. A nonparametric procedure for determining out of control lots is developed.

## 1 Introduction. Principles of SPC

Statistical Process Control (SPC) is based on the following assumptions about the production process:

- The characteristics of the output are normally distributed.
- When the process is running correctly (stays "in control"), the parameters of this distribution are equal to certain base values:  $\mu = \mu_0$ ,  $\Sigma = \Sigma_0$ .
- Malfunctions in the process lead to changes in value of one or both parameters  $\mu = \mu_1$ ,  $\Sigma = \Sigma_1$ . The process is then said to have gone "out of control".
- The process goes out of control as a result of well-defined, assignable and removable causes.

---

\*This research was sponsored by the ARO grant DAAL-03-91-G-0210

- Quality of the output is tied to its distributional properties. The items of good quality are those which come from the base distribution. The items produced while the process is out of control are of low quality.

These assumptions lead to the following basic paradigm of SPC:

1. Estimate the parameters of the base distribution of the process.

This is done based on the history of the process which is treated as a mixture of distributions containing the base and the contaminating, out-of-control distributions.

2. Knowing the base distribution, monitor the process to identify the non-conforming items.
3. Identify and remove the factors causing the non-conformities, and thus accomplish a lasting improvement in quality

The methods which can be used to implement SPC have been limited by the resources available in the industrial practice where trained statisticians are rare, most computing used to be done on hand-held calculators and the computations need to be done in the real time, since timely identification of the non-conforming lots is crucial for tracing the assignable causes. Those conditions resulted in an adherence to univariate methods and to the likelihood ratio tests based on the normal theory.

Recently, however, the picture has begun to change, as fast computers become cheaper and more popular. It is now possible to implement methods which are more computer intensive but also more robust and not confined to one dimension at a time. Below, we will present three such methods: a compound multivariate test to be used for identification of the out-of-control lots, a rank test for the same purpose and not relying on the assumption of normality, and the "King of the Mountain" procedure for estimating the mean of the base distribution.

## 2 Multivariate and Univariate Testing in SPC

Under the assumptions listed above, testing procedures for the new items are straightforward. To test for a shift in location, in a one dimensional situation, we use the likelihood ratio test based on statistic:

$$\frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma} \sim N(0, 1)$$

where  $\bar{X}$  is the lot average and  $n$  is the lot size. <sup>1</sup>

<sup>1</sup>The new items are not tested one by one but in batches which reduces the correlation between the consecutive test statistics and makes the assumption of normality more tenable

In the multivariate case this becomes:

$$n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}_0) \sim \chi_p^2$$

where  $p$  is the number of the process' characteristics, and we are using the  $\chi^2$  distribution since the actual variance-covariance matrix  $\boldsymbol{\Sigma}$  of the base distribution is assumed known. In both cases, the 0.002 significance level is usually used.

Despite their popularity with researchers in SPC, the multivariate techniques are not widely used in the industry practice. When confronted with a process with more than one quality characteristic, most professionals will treat it as a combination of several univariate processes, and test the new lots separately for each characteristic. This approach is perceived as conceptually and computationally simple, and also as facilitating discovery of Pareto glitches by identifying the out-of-control characteristic.

On the other hand, the dimension-by-dimension approach to multivariate cases has at least two obvious problems. Firstly, it misstates the significance level of the overall test. For if we compound  $p$  one-dimensional tests at 0.002 each, the resulting test will have the significance level not equal to 0.002 but ranging from 0.002 to  $1 - 0.998^p$  depending on the correlations among the variables.

Secondly, the "quasi-multivariate" test obtained by compounding univariate tests has a different rejection region than the  $\chi^2$  multivariate test. In the latter case, we have an ellipsoid which in the former case we try to approximate with a parallelepiped. As the dimensionality increases, one would expect the difference in their volumes to increase as well, and thus to cause a sizeable loss of power of the dimension-by-dimension test compared to the  $\chi^2$  test.

To verify this intuition, we investigated the powers of both kinds of tests against shifts in the mean vector of the distribution of the new lot. The power of the multivariate test was calculated as

$$P_p(\lambda) = \int_{(\frac{p+1}{p+2\lambda}) \chi^2_p(p)}^{\infty} d\chi^2(p + \frac{\lambda^2}{p+2\lambda})$$

where  $p$  is the dimensionality,  $\chi^2_p(k)$  is the 100  $(1 - \alpha)$  percentile of the  $\chi^2$  distribution with  $k$  degrees of freedom and  $\lambda$  is the noncentrality parameter:<sup>2</sup>

$$\lambda = n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}_0)$$

By analogy, the power of the univariate test is:

$$P_1(\lambda) = \int_{(\frac{1+\lambda}{1+2\lambda}) \chi^2_1(1)}^{\infty} d\chi^2(1 + \frac{\lambda^2}{1+2\lambda})$$

<sup>2</sup>A. Stuart, J. Ord, Kendall's Advanced Theory of Statistics, p. 870

where

$$\lambda = n \frac{(\bar{x} - \mu_0)^2}{\sigma}$$

Hence the power of the overall dimension-by-dimension test can be computed by taking the product:

$$P_T = 1 - \prod_{i=1}^p [1 - P_1(\lambda_i)]$$

where the noncentrality parameter is calculated separately for each dimension:

$$\lambda_i = n \frac{(\bar{x}_i - \mu_{0,i})^2}{\sigma_i}$$

To make a comparison possible, we adjusted the significance level of the dimension-by-dimension approach to:

$$\alpha^* = 1 - \sqrt[p]{1 - \alpha}$$

$\alpha$  was set at 0.002, the lot size at 10, and  $\Sigma$  was assumed equal to the identity.

The results of our inquiry are displayed in Figures 1—5. Figure 1 shows the powers of the  $\chi^2$  test and the dimension-by-dimension test for various slippage configurations. As we can see, in most cases, (about 75% of the points considered) the multivariate test is more powerful than the dimension-by-dimension test. There are cases however, (roughly 25% of the points considered) wherein most slippage occurs in a single dimension, and then the battery of univariate tests outperforms the  $\chi^2$  test. Consequently, we decided to do power comparisons for the two opposite slippage scenarios separately.

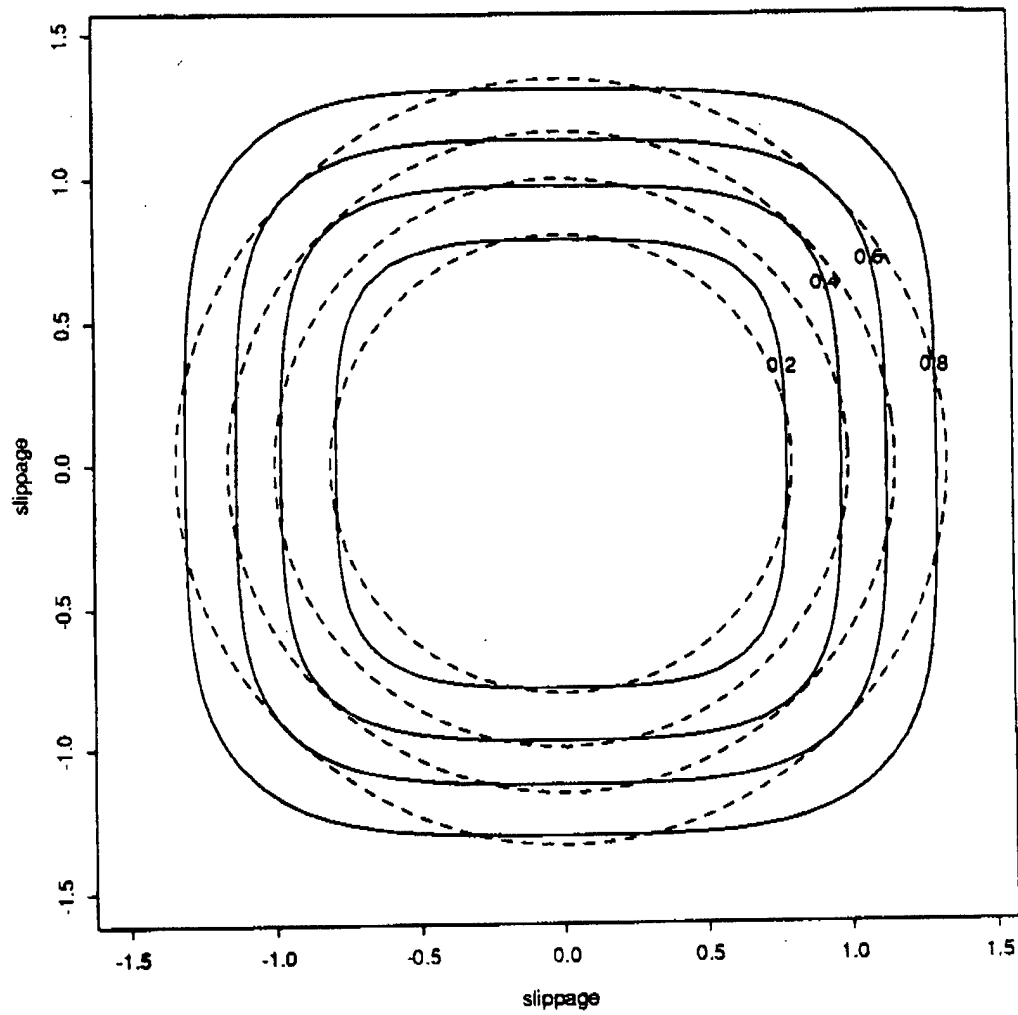
Figures 2, 3, and 4 show power curves and power ratios of both kinds of tests for 2, 5, and 10 dimensions. They indicate that the power loss resulting from using the multivariate test to detect the one-dimensional slippage is far less severe than the power loss resulting from using the dimension-by-dimension approach to detect a more balanced slippage configuration. Also, an increase in dimensionality clearly favors the multivariate test.

Furthermore, we have found that as the dimensionality increases, the percentage of cases in which the dimension-by-dimension test outperforms the multivariate test goes down, so that the overall performance of the  $\chi^2$  test is even better than one might infer from Figures 2—4.

As a caveat, one should remember that our study assumes an equal probability of all slippage configurations with the same noncentrality. In practice, this may not be the case. If there is evidence that a particular process goes out of control mostly one dimension at a time, the use of the dimension-by-dimension approach is much more justified than our results would indicate.

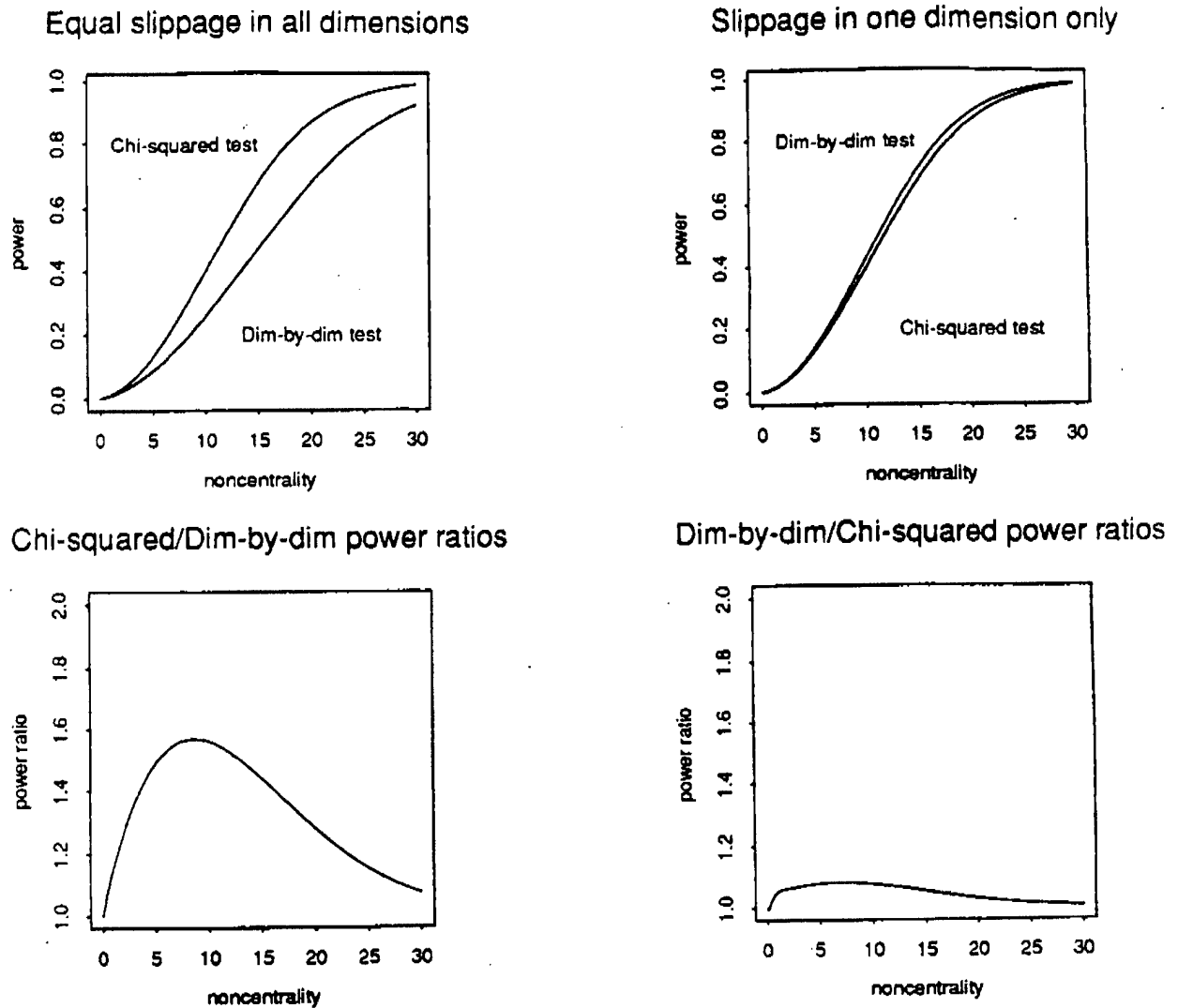
Finally, we considered performance of both tests when the characteristics of the process are correlated. One would expect the dimension-by-dimension approach to do poorly in this case, since looking at each characteristic separately

Figure 1.



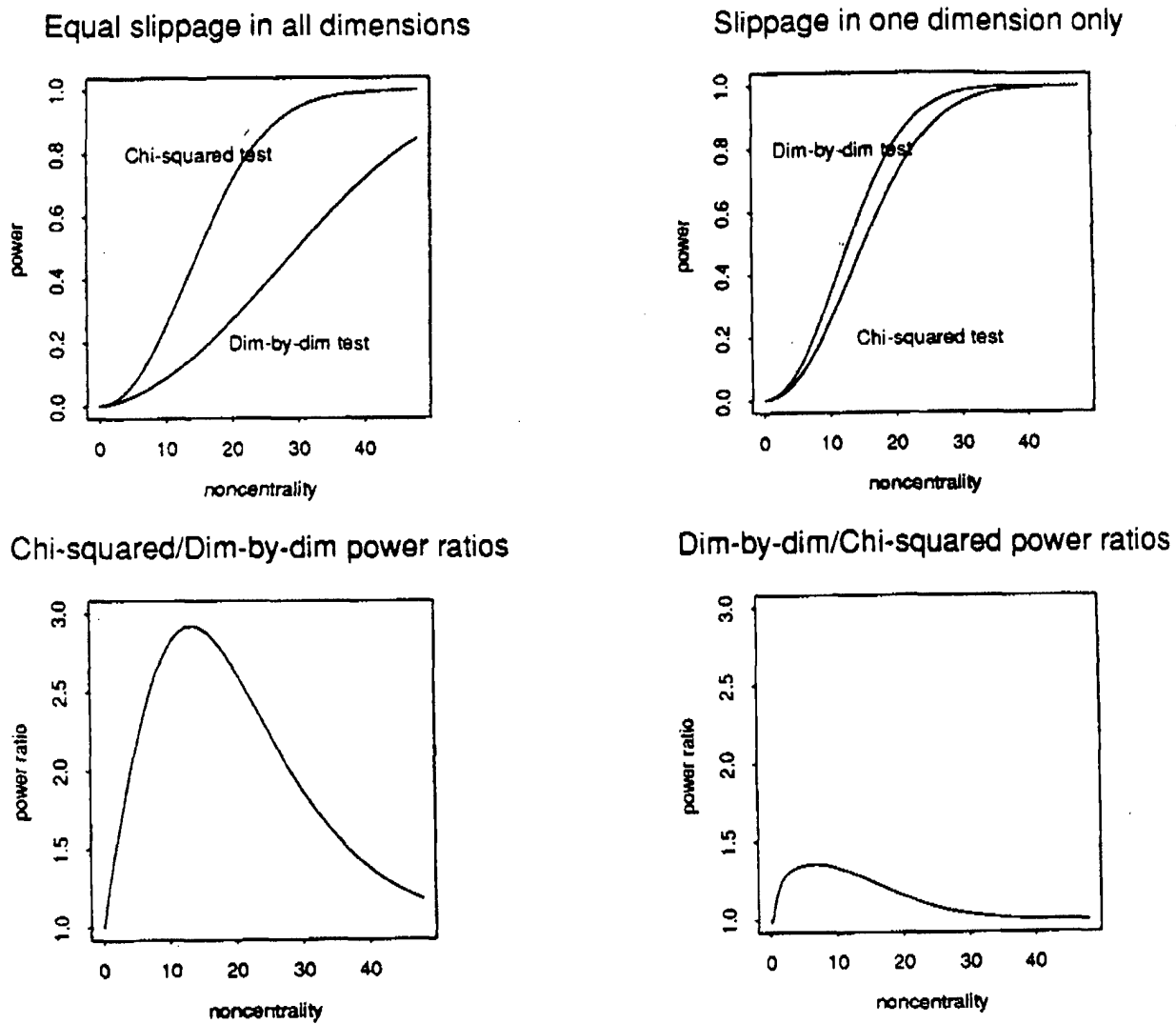
Comparison of power contours in two dimensions  
Chi-squared test: - - -      Dim-by-dim test: —

Figure 2.



Power curves and power ratios in two dimensions,  
under two slippage configurations

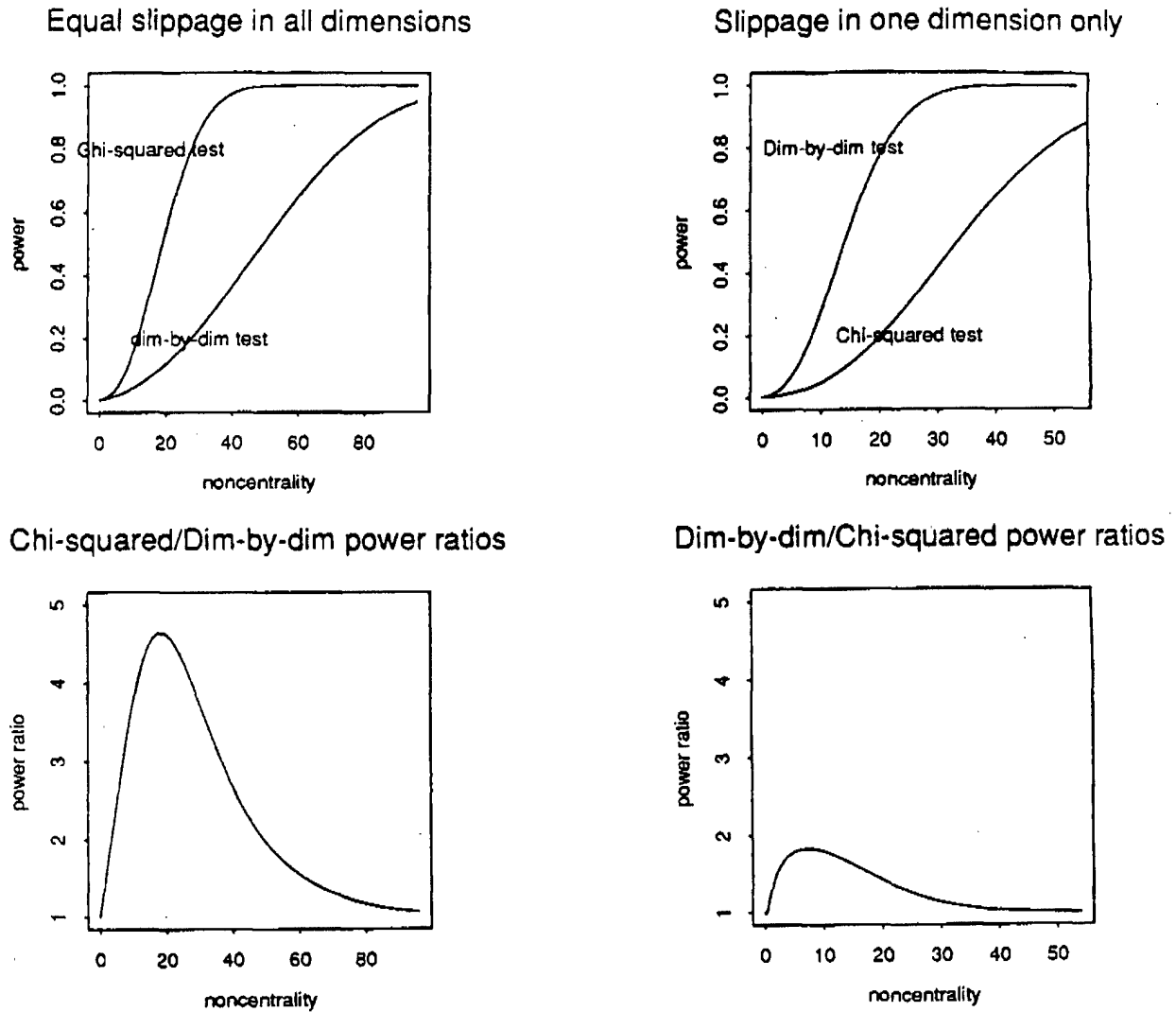
Figure 3.



Power curves and power ratios in five dimensions,  
under two slippage configurations

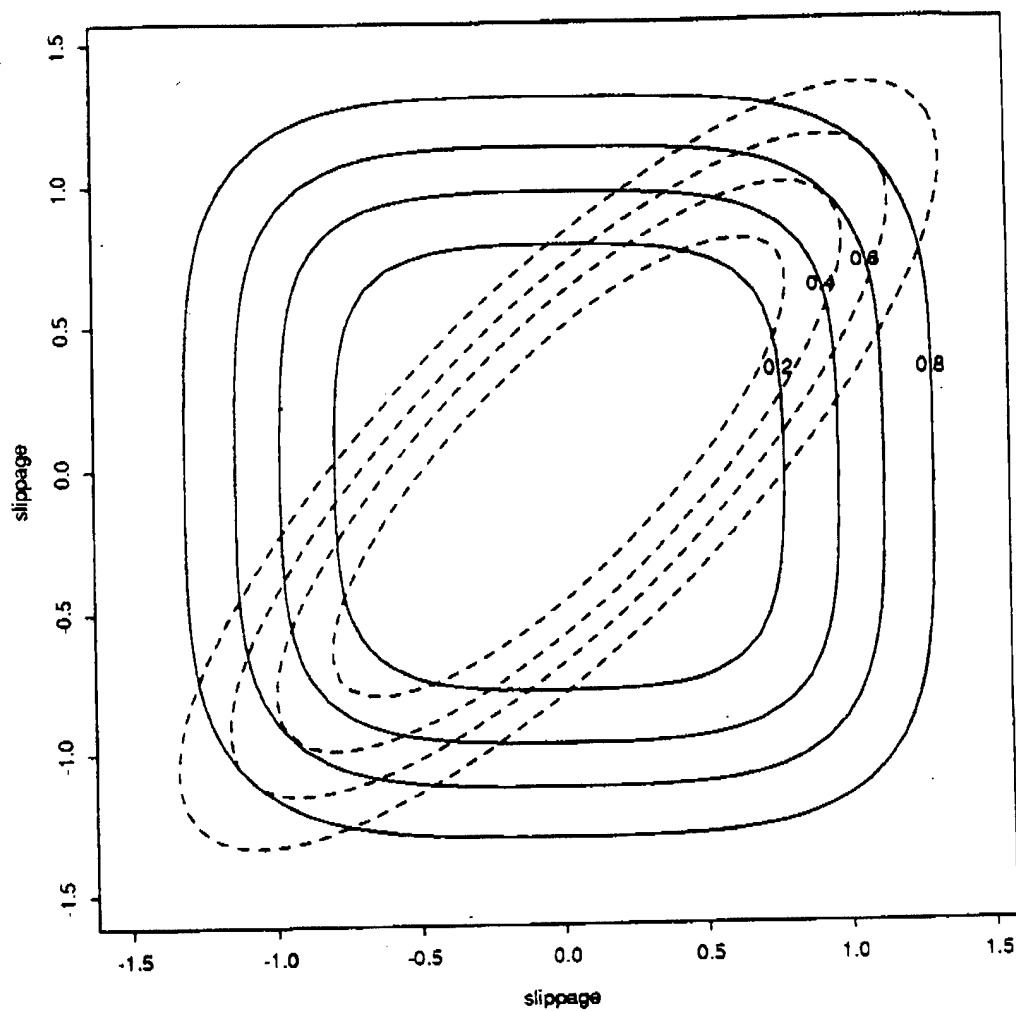


Figure 4.



Power curves and power ratios in ten dimensions,  
under two slippage configurations

Figure 5.



Comparison of power contours with correlation present  
Chi-squared test: - - - Dim-by-dim test: —

assumes their independence. Figure 5 shows an example of a bivariate process with the marginal variances equal to 1 and with the correlation coefficient  $\rho = 0.8$ . Clearly, testing one dimension at a time leads to a severe power loss for slippages in the direction opposite to the direction of correlation, but on the other hand, it actually gains us some power if the slippage is consistent with the covariance structure.

### 3 A Compound Test

Given the strong position of univariate tests in the SPC practice and the advantages they may have over the multivariate tests in certain cases, we do not expect the professionals to completely abandon the former for the latter. Therefore, we have investigated a possibility of combining both approaches in one compound test consisting of all possible univariate and multivariate tests performed on all the subsets of the characteristics under consideration.

Consider, for example, a process described by five characteristics. According to our compound procedure, we would first perform all five one-dimensional tests, each at the 0.002 significance level, followed by  $\binom{5}{2} = 10$  two-dimensional tests, again at 0.002 each, and then by  $\binom{5}{3} = 10$  three,  $\binom{5}{4} = 5$  four, and  $\binom{5}{5} = 1$  five-dimensional tests. On the whole, there will be  $(1+1)^5 - 1 = 31$  tests performed.

The advantage of this procedure lies in that it combines the good features of all the tests it consists of, and therefore, has the highest power for all alternative hypotheses, i.e., all slippage configurations.

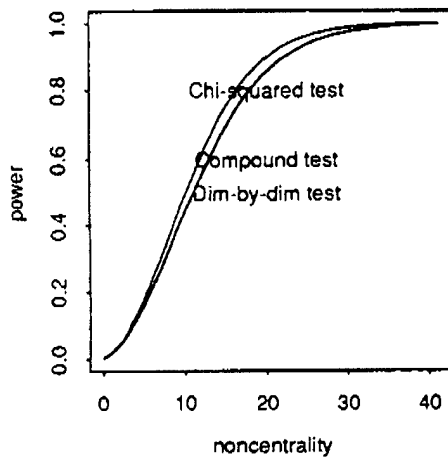
There is however, a price to be paid for this advantage. By compounding several tests into one, we are increasing the overall significance level. Using the Bonferroni inequality:

$$Pr\left(\bigcap_{i=1}^n A_i\right) \geq 1 - \sum_{i=1}^n Pr(A_i^c)$$

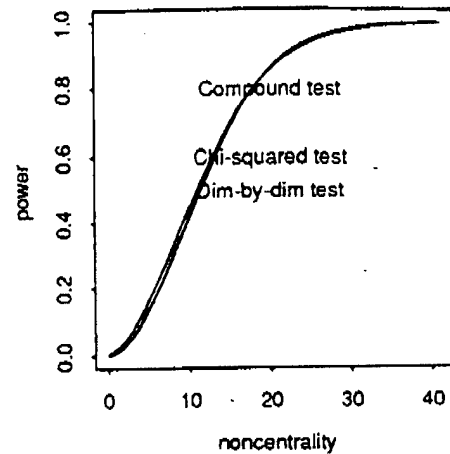
we can find an upper bound of this increase. For our example we obtain:  $\alpha^* = 31 * 0.002 = 0.064$ . This result, if precise, would disqualify the procedure as leading to a very large percentage of false positives. Fortunately, many of the individual tests are highly dependent, so we expect the actual significance level to be much smaller than the upper bound. The simulation results shown below confirm this expectation. In most cases, the overall  $\alpha$  is not much higher than the probability of the type I error resulting just from combining the one dimensional tests.

Figure 6.

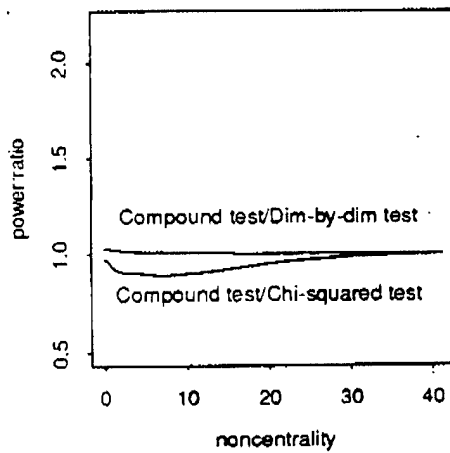
Alpha adjusted to the compound test



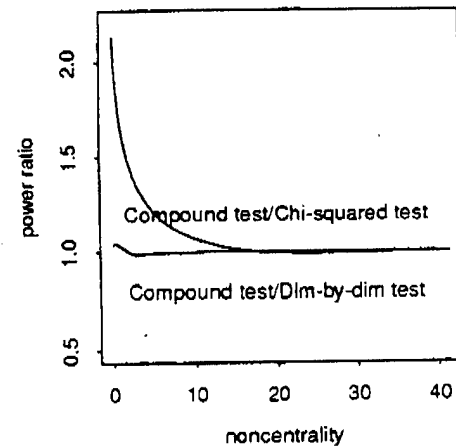
Component tests not adjusted



Power ratios



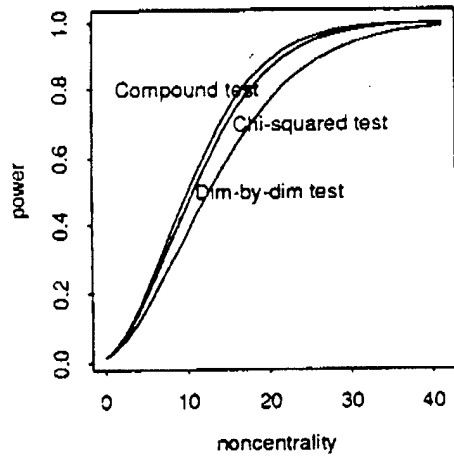
Power ratios



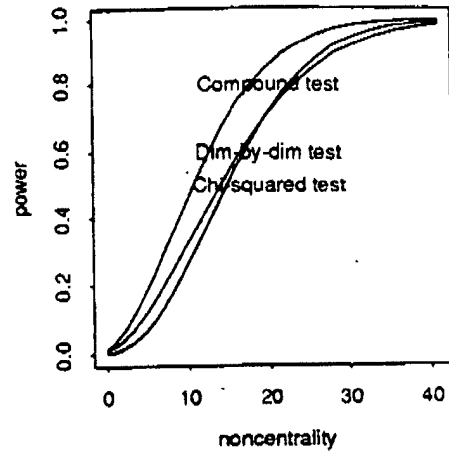
Power curves and smoothed power ratios of the compound test and its components in two dimensions under random slippage

Figure 7.

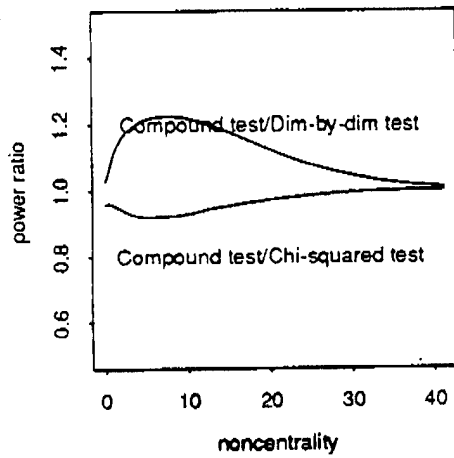
Alpha adjusted to the compound test



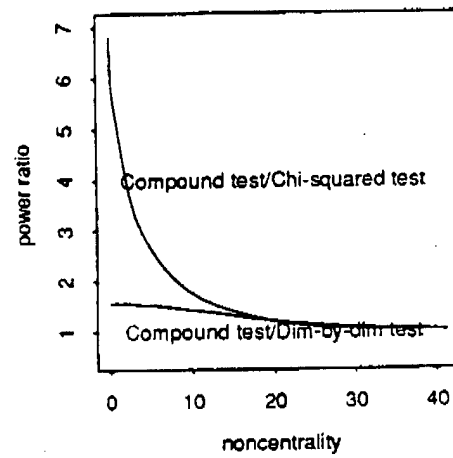
Component tests not adjusted



Power ratios



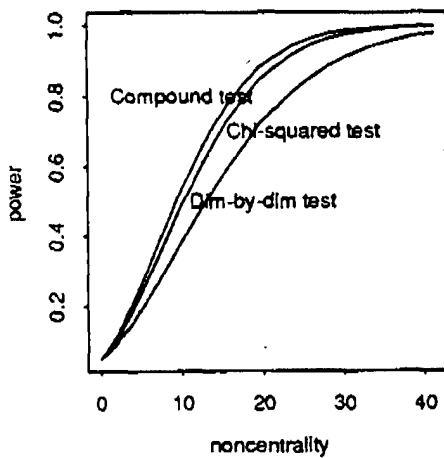
Power ratios



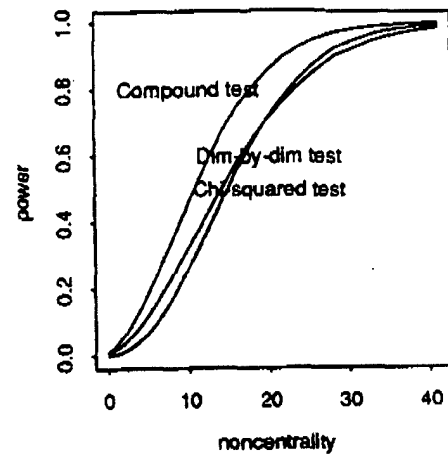
Power curves and smoothed power ratios in of the compound test and its components in five dimensions, under random slippage

Figure 8.

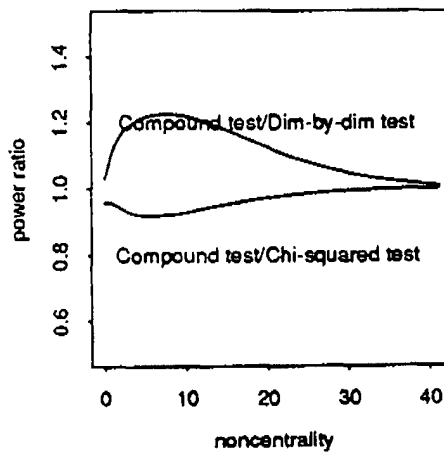
Alpha adjusted to the compound test



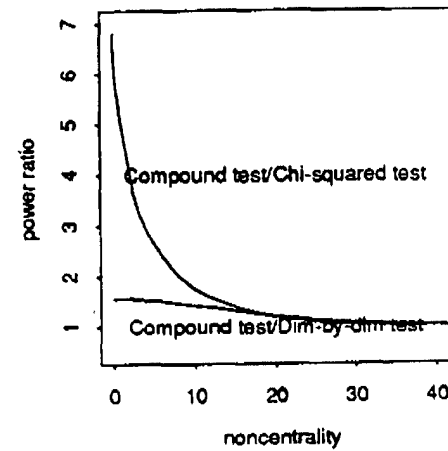
Component tests not adjusted



Power ratios



Power ratios



Power curves and smoothed power ratios of the compound test and its components in ten dimensions, under random slippage

Simulated  $\alpha$ -levels of the proposed compound test  
for various dimensionalities ( $p$ ) and lot sizes ( $n$ )

	n=5	n=10	n=15	n=20	n=50	n=100
p=2	0.00434	0.00466	0.00494	0.00508	0.00512	0.00528
p=3	0.00756	0.00722	0.00720	0.00720	0.00704	0.00826
p=4	0.01126	0.01072	0.01098	0.01098	0.01108	0.01136
p=5	0.01536	0.01482	0.01552	0.01544	0.01706	0.01670
p=10	0.0446	0.04674	0.04474	0.04468	0.04440	0.04674

Figures 6—8 illustrate a comparison of the proposed compound test to its natural competitors. For dimensions  $p = 2$ ,  $p = 5$ , and  $p = 10$ , we compared the power of the compound test to the powers of its one- and  $p$ -dimensional component tests, as well as to the powers of those tests when adjusted for the increase of the significance level. All results are based on 50,000 simulations, and on the assumption that the shifted lot means are distributed uniformly on the surface of the hypersphere with radius equal to  $\sqrt{\frac{\lambda}{n}}$ .

Given the results obtained, we think that the tradeoff between the increase in significance level and the increase in power is advantageous in this case and warrants implementation of the procedure.

#### 4 A Rank Test for Slippage in Location

The assumption of normality is accepted in the theory and the practice of SPC, and it seems to be working well. Still, making this assumption may be very hard to justify in some cases, and then, a distribution-free tests would provide an interesting alternative.

We would like to suggest a simple, distance-based rank test for shifts in location. It can be characterized as follows:

1. Let:

- $X_1, \dots, X_N$  be the mean vectors of the base lots
- $Y$  be the mean of the new lot
- $D_{i,j} = (X_i - X_j)' (X_i - X_j)$
- $Q_i = (X_i - Y)' (X_i - Y)$
- $D_{i\cdot} = \frac{1}{N} \sum_{j=1}^N D_{i,j}$
- $Q = \frac{1}{N} \sum_{j=1}^N Q_j$

2. The test rejects the hypothesis of no shift in location if

$$Q > \text{Max}(D_1, \dots, D_N)$$

3. The significance level of this test is  $\frac{1}{N+1}$

To establish the validity of this test we did two stages of simulation studies. First, we investigated the relative performance of this test and of the likelihood ratio test in the case in which the data was indeed normal and the slippage was the same in all dimensions. Under these conditions, the likelihood ratio test has more power than any other test of its generality, so we were not hoping to improve on it, but only to show that the use of our rank test would not lead to a catastrophe.

Figures 9 and 10 display our results based on 20,000 simulations with  $N=50$ . Figure 9 shows the power curves of the rank test for dimensions 2—5 and for the diagonal covariance matrices both in the base and in the new lot. Figure 10 shows the corresponding power ratios, where the denominator is based on the likelihood ratio  $\chi^2$  test with the same significance level as the rank test. We have observed a rather poor performance for 2 dimensions, which, however, improves fast with dimensionality and becomes quite good at  $p = 5$

In the second stage, we replaced the normal distribution from the first stage with the multivariate T-distribution with 3 degrees of freedom. Notably if

$$\begin{aligned} Z &\sim MVN(0, \Sigma) \\ V &\sim \chi^2_{(p)} \\ X &= \frac{Z}{\sqrt{V/p}} + \mu \end{aligned}$$

then

$$X \sim T_p(\mu, \Sigma)$$

As expected, the rank test performed worse for small slippages in this case, due to the tailiness of the  $T_3$ . It worked well for larger slippages and also maintained its significance level. The results are shown in Figure 11.

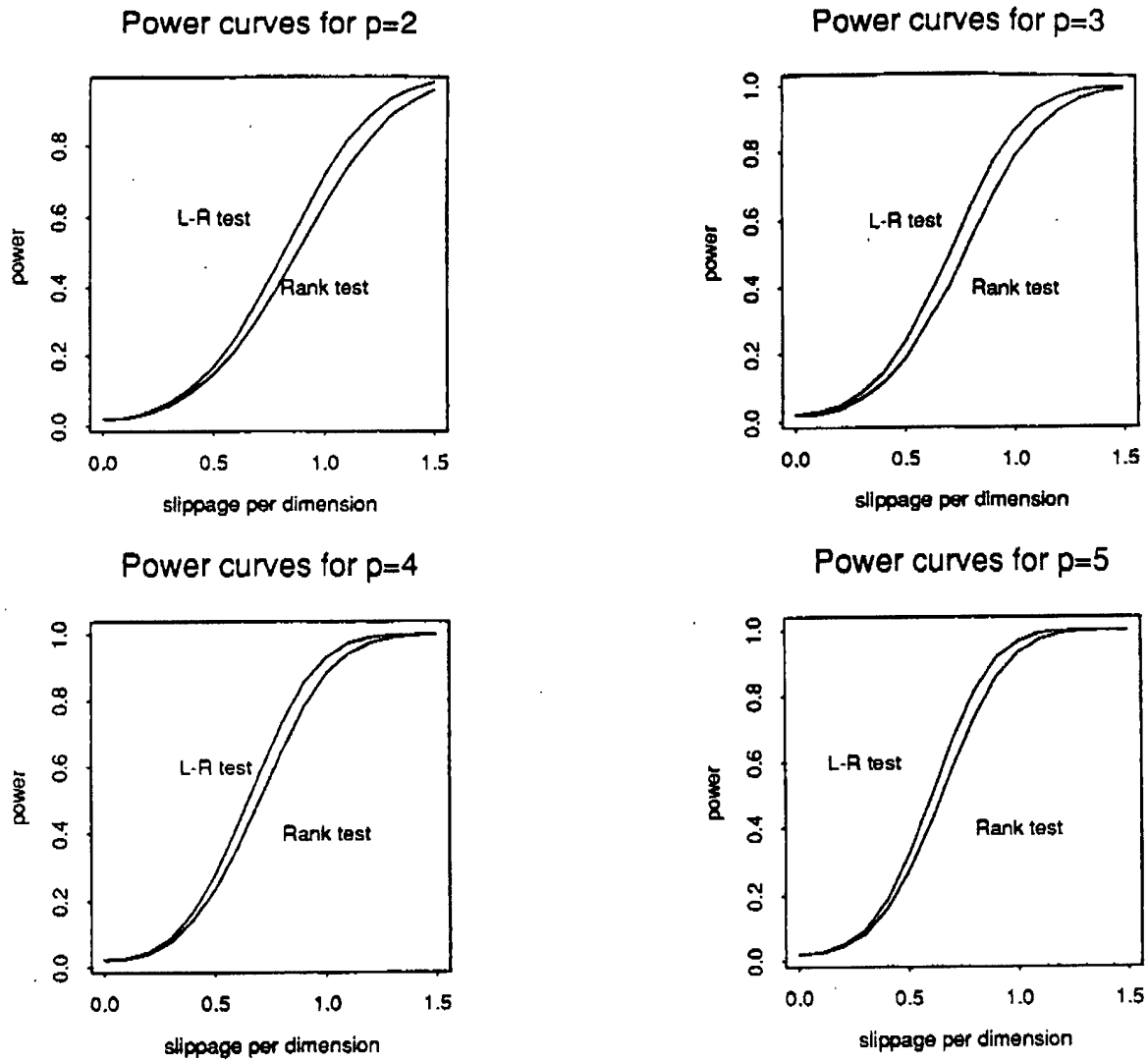
On the other hand, the likelihood ratio test failed completely in this case, due to a dramatic increase in the significance level:

$p = 2$	-	$\alpha = 0.12705$
$p = 3$	-	$\alpha = 0.25765$
$p = 3$	-	$\alpha = 0.29925$
$p = 5$	-	$\alpha = 0.3132$
$p = 10$	-	$\alpha = 0.55595$

the intended significance level was  $\alpha = 0.02$  in each case.

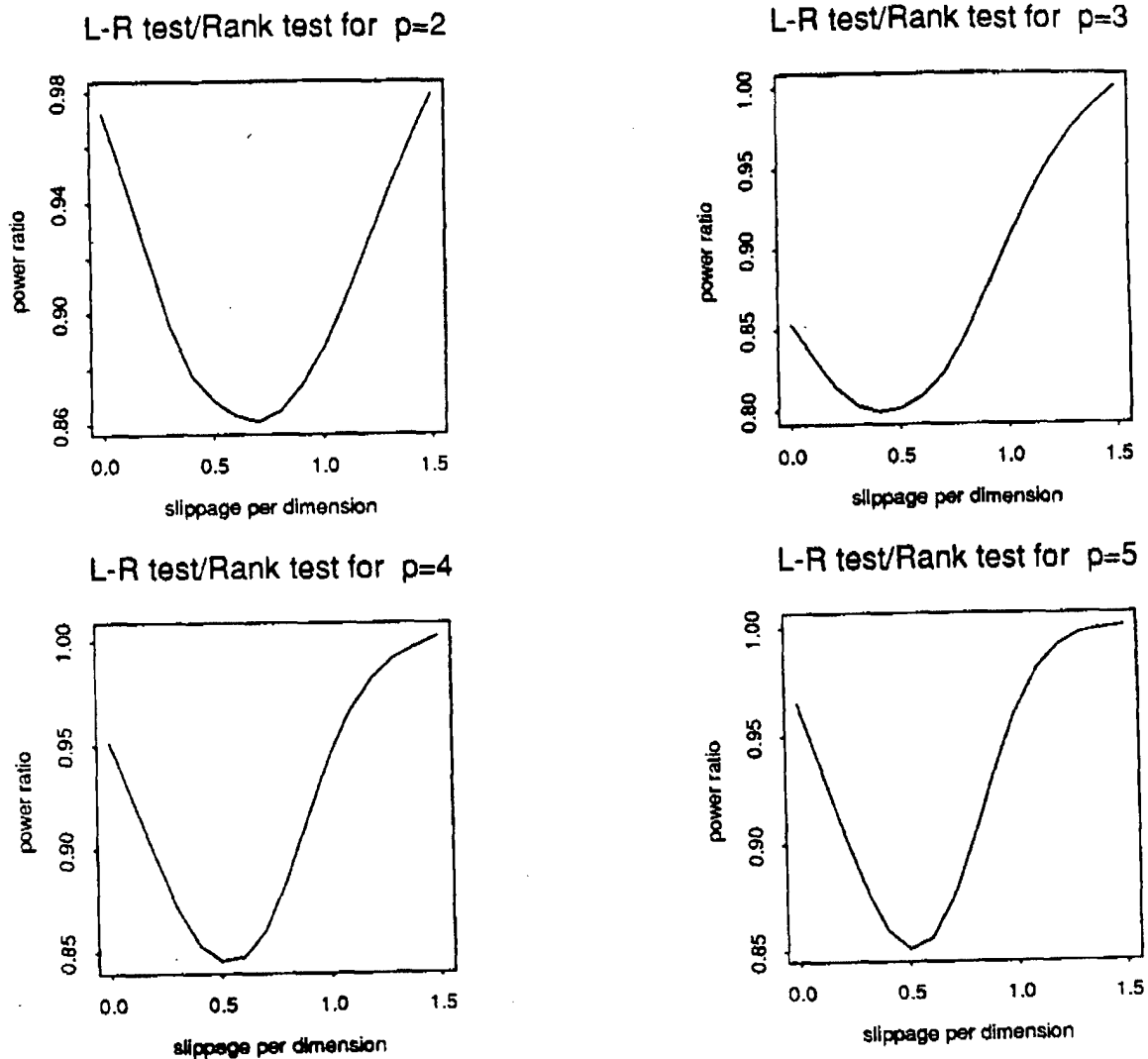


Figure 9.



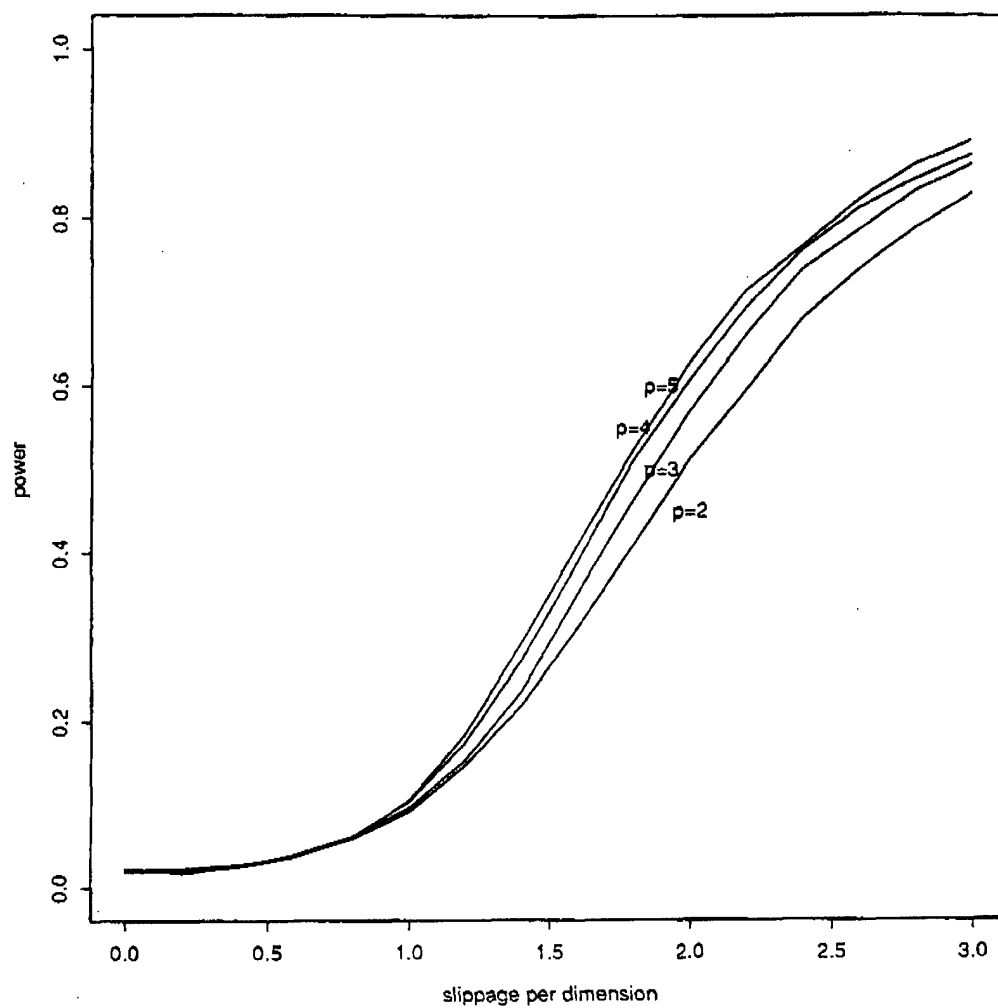
Power curves of the rank test and the L-R test  
under equal slippage in all dimensions

Figure 10.



Smoothed power ratios in of the L-R and the rank tests  
under equal slippage in all dimensions

Figure 11.



Power curves of the first rank test for T-3,  
under equal slippage in all dimensions

## 5 A Portmanteau Rank Test

The test described above is simple and works well in many cases. Another procedure, we have developed, shows a better performance at the price of higher complexity. It can be described as follows:

1. Let

- $\{X_{i,j} : i = 1, \dots, n, j = 1, \dots, N\}$  be  $N$  lots of  $n$  items from the base distribution
- $\{Y_i : i = 1, \dots, n\}$  be the new lot
- $\bar{X}_i$  be the mean vector of the  $i$ th base lot
- $\bar{\bar{X}}$  be the mean of the lot means
- $S_i$  be the sample covariance matrix of the  $i$ th base lot
- $\bar{S}$  be the elementwise average of the sample covariance matrices of the base lots
- $Z_{i,j}$  be the  $X_{i,j}$  transformed as:  $Z_{i,j} = \bar{S}^{-1/2} (X_{i,j} - \bar{\bar{X}})$
- $\bar{Z}_i$  be the mean vector of the  $i$ th transformed base lot
- $G_i$  be the sample covariance matrix of the  $i$ th transformed base lot
- $Q_i$  be the  $Y_i$  transformed by  $Z$ :  $Q_i = \bar{S}^{-1/2} (Y_i - \bar{\bar{X}})$
- $\bar{Q}_i$  be the mean vector of the transformed new lot
- $M$  be the sample covariance matrix of the transformed new lot

2. For each  $\bar{Z}_i$  calculate

$$\|\bar{Z}_i\| = \sum_{j=1}^p \bar{Z}_{i,j}^2$$

3. For each  $G_i$  calculate:

$$\|G_i\| = \sum_{j=1, k=1}^p G_{i,j,k}^2$$

4. Analogously, calculate  $\|\bar{Q}\|$  and  $\|M\|$

5. The test rejects the null hypothesis of the new lot coming from a distribution similar to the base if:

$$\|M\| > \text{Max}(\|G_1\|, \dots, \|G_N\|) \quad \text{and} \quad \|\bar{Q}\| > \text{Max}(\|\bar{Z}_1\|, \dots, \|\bar{Z}_N\|)$$

6. Since the probability of a type-I error for each element of the conjunction is  $\frac{1}{N+1}$ , the overall significance level of this test is approximately

$$1 - \left[1 - \frac{1}{N+1}\right]^2$$

Clearly, this procedure is much more complex and computer intensive than the previous one. To see whether the price is worth paying, we studied the performance of this test in the same way as previously. First then, we compared our procedure to the likelihood ratio test in the case in which the data was normal and the slippage was the same in all dimensions. Secondly, we looked at the  $T_3$  case. Since there was no change in the covariance structure of the new lot, we looked only at the location-based part of the test.

The results for the normal case are summarized in Figures 12 and 13 which display the power curves and the corresponding power ratios. As above, we have observed a rapid improvement of performance with the increase in dimensionality.

The results for  $T_3$  are shown in Figure 14. There is a sizeable gain in power as compared to the first rank test, which we attribute to the second test's better handling of outliers.

Better performance is not the only advantage of this test however. A robust standardization of the base and the new lots will prevent spurious results to which the first rank test is susceptible if the scales of different variables are different and/or if the variables are correlated. For suppose that  $X_1$  has the marginal variance twice as large as  $X_2$  and that the slippage occurs in the direction of  $X_2$  only. The first test is likely to fail to reject under these circumstances, because the average distance to the new mean vector, although large compared to the variability of  $X_2$ , will nevertheless be small compared to base distances along  $X_1$ .

This problem could be remedied within the framework of the first test by expressing each variable in units of its marginal variance. It can reappear though if the base data are correlated, because, just as in the case of unequal marginal variances, the slippage contrary to the direction of correlation would pass unnoticed, whereas the slippage in that direction would be overemphasized.

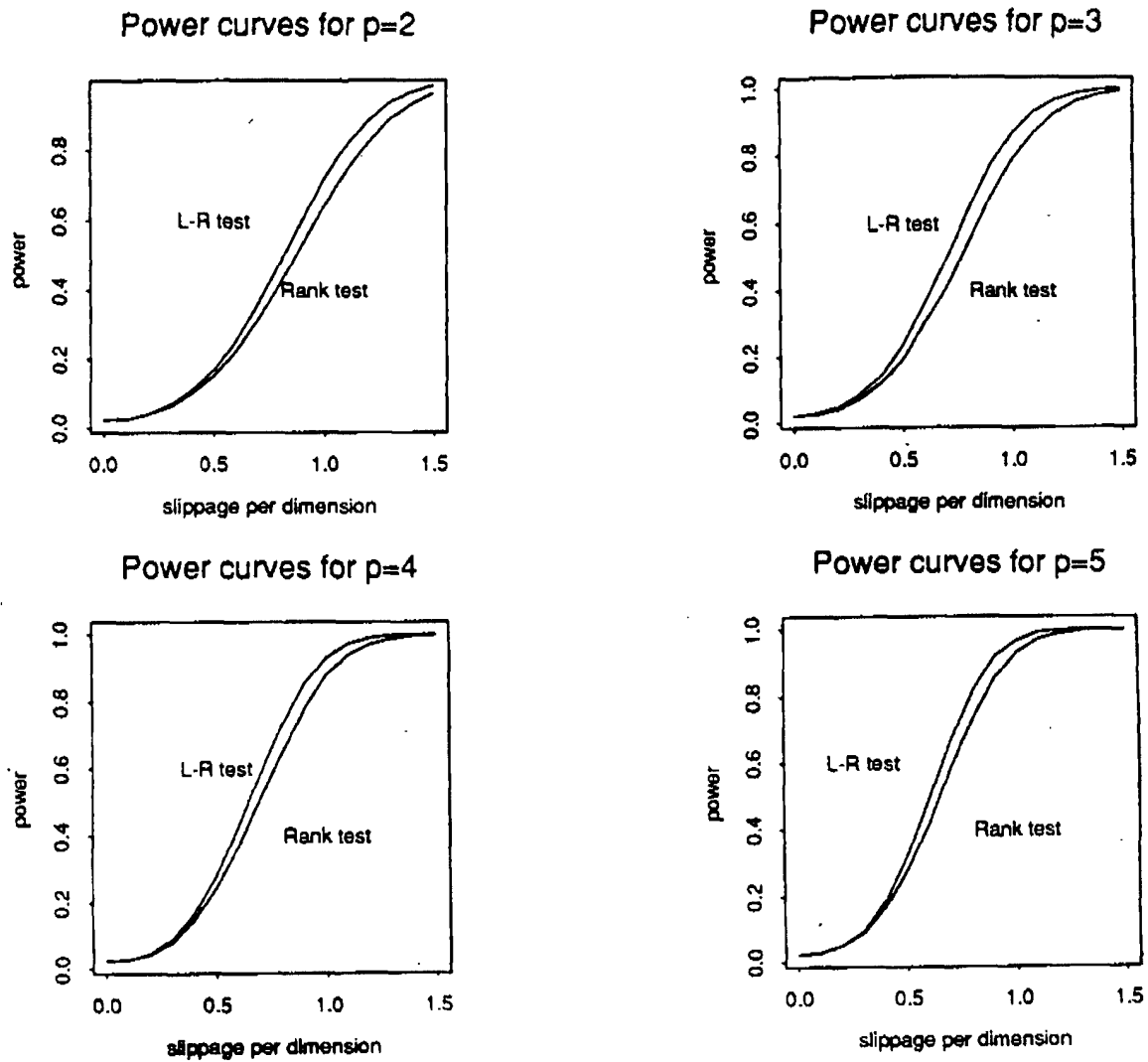
The second test does not seem to have this problem.

Finally, the second test is more versatile than the first as it can detect changes in the covariance structure of the new lots and not only shifts in location.

## 6 Estimating the Mean of the Base Distribution

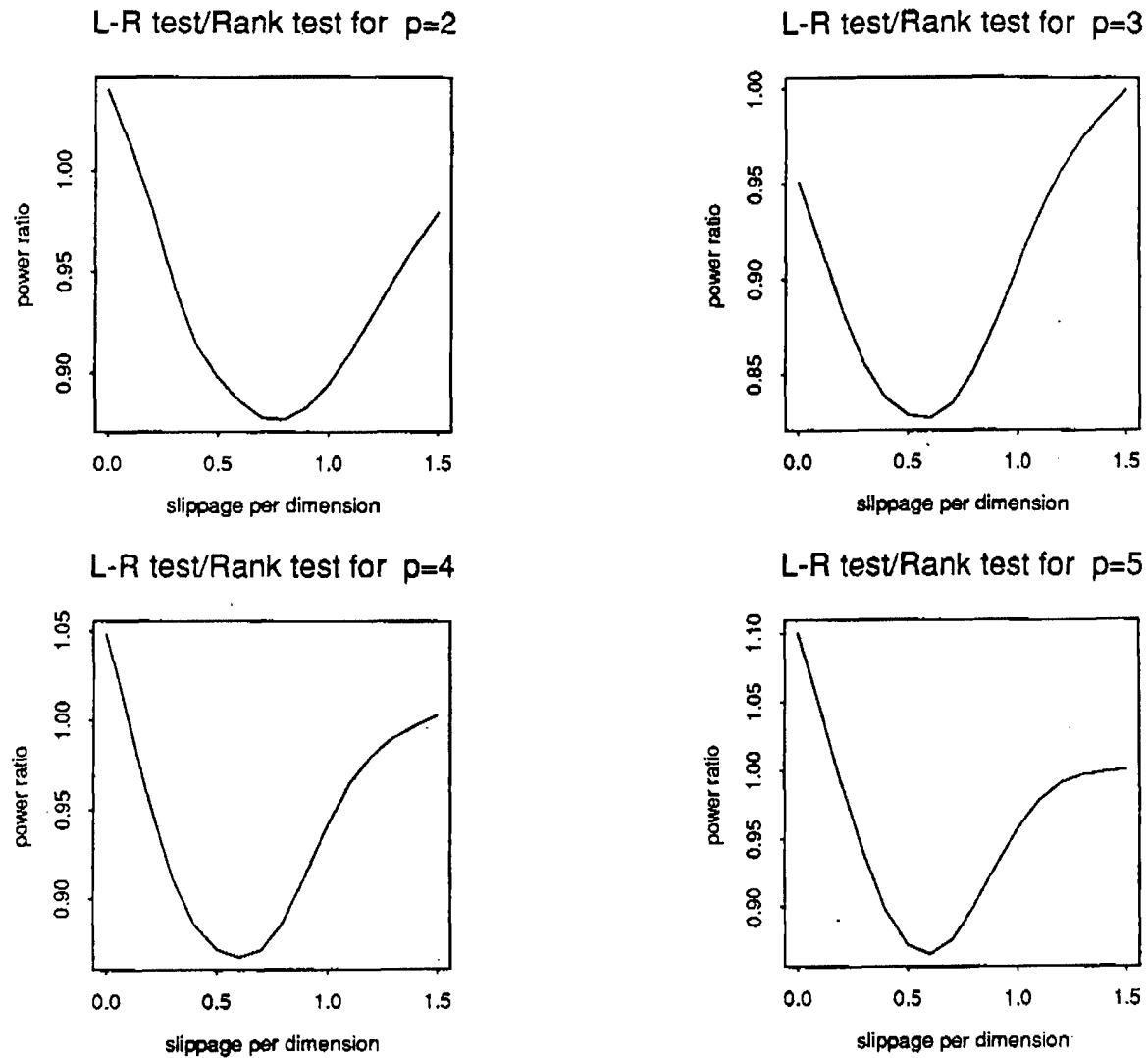
The estimation of the parameters of the base distribution, which must precede testing of the new items, is often presented as a testing procedure itself. It is

Figure 12.



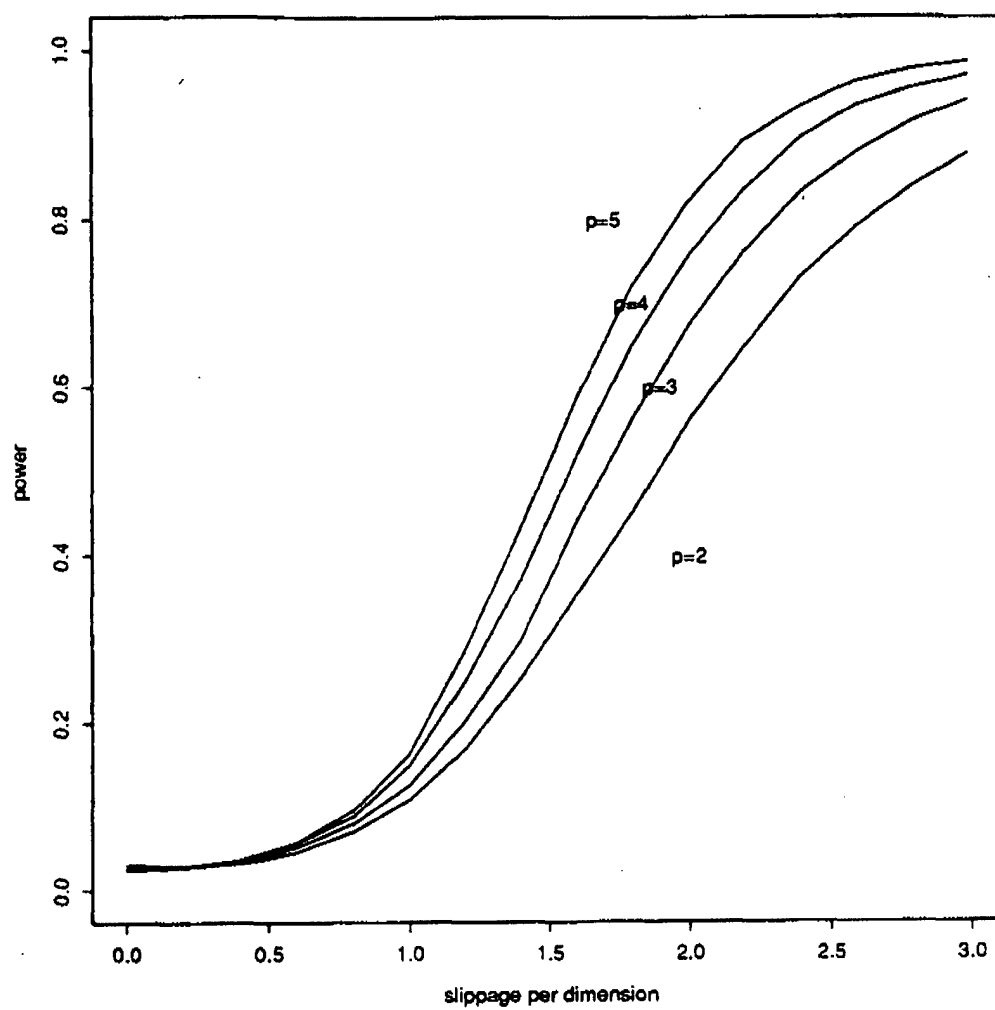
Power curves of the second rank test and the L-R test  
under equal slippage in all dimensions

Figure 13.



Smoothed power ratios in of the L-R and the second rank tests  
under equal slippage in all dimensions

Figure 14.



Power curves of the second rank test for T-3,  
under equal slippage in all dimensions



said to consist in taking the history of the process as the sample, performing a LR test on each point to see whether it conforms to the base distribution, and in removing the non-conforming points. The estimate of the base location parameter is then the mean vector of the points left in the sample.

Since the statistics required for the test:  $\bar{x}$  and  $\Sigma$  are now obtained from the very same points being tested, one must have the procedure iterate: first, one will take the overall mean of the reference points as  $\bar{x}$  and the mean of lot variances as  $\Sigma$ , then perform a likelihood ratio test on each reference point and exclude the points for which the test rejects. In the next iteration, the mean and the variance are calculated only from the points left in the sample by the previous one. The procedure will continue until no more points are rejected.

Despite such an adjustment, the procedure is not valid. For the likelihood ratio tests, it involves, would be valid only if the base distribution had the same mean and variance as the joint distribution of all reference points. This is true, however, only with no contamination, but then we do not need to test at all. Indeed, if there is contamination present, the reference sample mean may have a completely unreasonable value, especially if the mean(s) of the contaminating distribution(s) is far from the base mean.

To neutralize this problem, it suffices to notice that the procedure described above is not really a test but an iterative trimming. The likelihood ratio rejection rules do not have any probabilistic interpretation here. They constitute only a stopping rule, chosen more or less arbitrarily, perhaps with consideration of the fact that, if indeed there is no contamination in the base sample, our procedure will reject only few points if any.

This procedure is more robust than the simple  $\bar{x}$  because of the use of trimming, but still shares some of its flaws: it is likely to be suboptimal if the proportion of contaminants in the sample is high, and especially if their distribution is well separated from the base. Consequently, it offers the least protection against the cases which are the most serious, and, intuitively, the easiest to handle.

Other robust estimators of location have been proposed as an improvement over the trimmed mean, and they seem to be working well in one dimension. To use them in the multivariate settings is more difficult, however, because there is no natural, unique ordering here to rely upon. Depending on the approach, there may be several multivariate generalizations of the same univariate quantity. For the most popular robust estimator of location, the median, we have at least two generalizations:

- the arithmetic median is the vector of the medians of individual variables. It is based on the definition of the median as the 50th percentile.
- the geometric median is defined as the point such that the sum of its distances from the sample points is a minimum. It is based on the definition:

$$M = \arg \min \int_{-\infty}^{\infty} |x - y| dF(x)$$

The geometric median is invariant under the rotation of axes, which is what one would expect from a point "central" to the data set. Its only drawback is its scale-dependence which may create problems if the contaminating distribution has the covariance structure different than the base.

The arithmetic median has an advantage of being scale invariant but its dimension-by-dimension approach is more likely to fail for small separations between the base and the contamination, in which case the contaminating points may stick out of the main cluster without actually sticking out in any particular dimension.

The following tables illustrate the behavior of both estimators. Displayed are the mean squared errors per dimension for the estimates of the mean of the dominant distribution, in dimensionalities 2, 5, and 10, mixing proportions of 60%, 75% and 90% of observations coming from the base distribution and with separations between the base and the contaminations of 1, 3 and 5. The results are based on 100 simulations in each case, with the number of contaminating distributions varying randomly between 1 and 5.

#### The Arithmetic Median

**p=2**

n=50				n=100		
dist	60%	75%	90%	60%	75%	90%
1	0.0655	0.0406	0.0350	0.0463	0.0282	0.0166
3	0.2318	0.1121	0.0382	0.2025	0.0709	0.0224
5	0.3251	0.1186	0.0437	0.3232	0.0857	0.0249

**p=5**

n=50				n=100		
dist	60%	75%	90%	60%	75%	90%
1	0.0443	0.0391	0.0298	0.0292	0.0206	0.0171
3	0.1315	0.0704	0.0348	0.1254	0.0512	0.0206
5	0.2298	0.1001	0.0428	0.1894	0.0677	0.0212

**p=10**

n=50				n=100		
dist	60%	75%	90%	60%	75%	90%
1	0.0390	0.0355	0.0328	0.0238	0.0192	0.0163
3	0.0893	0.0518	0.0390	0.0783	0.0352	0.0178
5	0.1559	0.0777	0.0369	0.1367	0.0480	0.0213

### The Geometric Median

$p=2$

n=50				n=100		
dist	60%	75%	90%	60%	75%	90%
1	0.0603	0.0377	0.0288	0.0424	0.0280	0.0138
3	0.2287	0.0953	0.0317	0.1934	0.0637	0.0209
5	0.2926	0.0967	0.0347	0.3017	0.0785	0.0226

$p=5$

n=50				n=100		
dist	60%	75%	90%	60%	75%	90%
1	0.0347	0.0284	0.0208	0.0269	0.0170	0.0120
3	0.1167	0.0537	0.0249	0.1166	0.0438	0.0148
5	0.2108	0.0838	0.0294	0.1769	0.0611	0.0166

$p=10$

n=50				n=100		
dist	60%	75%	90%	60%	75%	90%
1	0.0282	0.0246	0.0217	0.0269	0.0170	0.0120
3	0.0733	0.0411	0.0245	0.1166	0.0438	0.0148
5	0.1383	0.0613	0.0251	0.1769	0.0611	0.0166

The corresponding results for the trimmed mean are:

$p=2$

n=50				n=100		
dist	60%	75%	90%	60%	75%	90%
1	0.0566	0.0345	0.0250	0.0426	0.0266	0.0118
3	0.3433	0.1644	0.0431	0.3100	0.1404	0.0314
5	0.8724	0.4099	0.0843	0.9458	0.3377	0.0668

$p=5$

n=50				n=100		
dist	60%	75%	90%	60%	75%	90%
1	0.0329	0.0276	0.0191	0.0272	0.0159	0.0106
3	0.1426	0.0716	0.0272	0.1524	0.0637	0.0184
5	0.3929	0.1992	0.0458	0.3564	0.1628	0.0315

p=10

dist	n=50			n=100		
	60%	75%	90%	60%	75%	90%
1	0.0279	0.0235	0.0208	0.0176	0.0136	0.0106
3	0.0829	0.0489	0.0248	0.0835	0.0366	0.0141
5	0.2057	0.1022	0.0315	0.1966	0.0796	0.0218

The median-based estimators are less sensitive to the outliers than the trimmed mean and handle better the high contamination/large separation cases. The trimmed mean is superior in "well-behaved" cases with a small separation and a high proportion of "in-control" observations.

## 7 "King of the Mountain" Algorithm

In an attempt to improve on the estimators presented above, we tried to identify and alleviate the shortcomings of each. The median-based estimators show high variability and a limited ability to handle the high-separation cases, which seems due to the fact that they only discount the outlying observations rather than remove them from the sample.

The trimmed mean has a lower variability but can prove disastrous for the high-separation cases. Both features are rooted in its low selectivity: in our simulation studies no more than 6%-8% of observations were removed from the sample, even under severe contamination. Naturally, the overall proportion of rejected points can be increased by diminishing the "acceptance region" on which the trimming is based. This however, leads to a biased estimate because too many base points are thrown away.

Based on those observations, we suggest the following "King of the Mountain" algorithm:

1. Find  $\bar{\bar{X}}$ , the mean of the means of the base lots
2. Find two lots whose means are the furthest apart
3. Remove this of the two lots found in (2) whose mean further from  $\bar{\bar{X}}$
4. Repeat 1—3 until the number of lots removed equals the expected number of contaminated lots in the base sample

This procedure assumes that component distributions of the mixture are spherical (such as  $MVN(\mu, I)$ ), but the distances among their means are large enough to make the mixture asymmetrical. The algorithm looks at the extreme distances among sample points to identify the directions in which the asymmetries occur and to trim the sample along those directions. In an attempt to retain the base points while removing the contaminations, it focuses on one

"side" of the current mean, and throws away this of the extreme points which is further from the mean.

We have done simulation studies of the algorithm's performance in the same way as for the other estimators. The results are summarized in the following tables:

$p=2$

n=50				n=100		
dist	60%	75%	90%	60%	75%	90%
1	0.1047	0.0593	0.0306	0.0665	0.0393	0.0144
3	0.1503	0.0536	0.0255	0.0660	0.0265	0.0156
5	0.0528	0.0286	0.0269	0.0178	0.0146	0.0105

$p=5$

n=50				n=100		
dist	60%	75%	90%	60%	75%	90%
1	0.0612	0.0364	0.0236	0.0366	0.0236	0.0131
3	0.1056	0.0488	0.0248	0.0558	0.0263	0.0132
5	0.0565	0.0309	0.0250	0.0215	0.0154	0.0103

$p=10$

n=50				n=100		
dist	60%	75%	90%	60%	75%	90%
1	0.0473	0.0352	0.0241	0.0294	0.0194	0.0119
3	0.0693	0.0389	0.0249	0.0559	0.0230	0.0129
5	0.0407	0.0304	0.0237	0.0243	0.0148	0.0113

As the simulations indicate, "King of the Mountain" performs well even for small separations and relatively high proportion of contaminating lots in the base sample, although it underperforms the trimmed mean in those cases since the asymmetries in the sample are not large enough. The algorithm's performance improves with the magnitude of the slippage and the increase in dimensionality. Notably, it is the only one among the estimators studied whose performance improves as the separation of component distributions increases.

## 8 Conclusions

Wide availability of fast computers may revolutionize SPC, allowing its professionals to use more sophisticated, more computationally intense procedures such as multivariate and nonparametric techniques.

## **Sampling Problems Pertaining to the Number of Replications for Stochastic Simulation Models**

**William E. Baker**

**David W. Webb**

**Lawrence D. Losie**

**Army Research Laboratory  
APG, MD 21005-5068**

The Survivability/Lethality Analysis Directorate of the Army Research Laboratory utilizes a hierarchy of simulation models to evaluate the vulnerability of armored fighting vehicles. Central to this process is the examination of the damage state of critical components. The damage state vector is actually an  $n$ -tuple, each element of which represents a critical component. Thus, each element can take on the value 0 (no kill) or 1 (kill), implying that for a system with  $n$  critical components, there are  $2^n$  possible damage states. There is interest in the distribution of this random variable. In attempting to compare the consistency of live-fire results with the output from a stochastic simulation model, the following problems have arisen concerning the distribution of component damage states:

- 1) There are instances when the live-fire result does not match the output from any individual replication of the simulation.
- 2) There are instances when many (much greater than 5%) of the replications of the simulation output a unique damage state, thus making the tail of the empirical distribution function rather nebulous.

Assuming we know the probability of a kill ( $p$ ) for each of  $n$  independent components in a damage state, we have been able to determine how many shots must be fired (or, alternatively, how many times the simulation must be replicated) to be  $x\%$  confident that we have seen  $y\%$  of the distribution of the damage states. However, for many values of  $p$  and  $n$  and typical values of  $x$  and  $y$ , this number of shots/replications is impracticable. The question then follows:

- ◆ Are there statistical techniques that allow us to address the problems mentioned above when the number of simulation replications necessary to adequately describe the distribution of the damage state vectors is impracticable?

## Introduction to Question

Over the years the Ballistic Research Laboratory (now the Army Research Laboratory) has utilized a hierarchy of simulation models to evaluate the vulnerability of armored fighting vehicles. This hierarchy includes a stochastic model which was developed in the late 1980's. In attempting to check the consistency of results from this model with those of live-fire tests, we have developed a statistical procedure for which the null hypothesis is: *"Results from the live-fire tests are consistent with output from the simulation model."* This procedure concentrates on the damage state vectors.

A damage state vector represents the damage state of the vehicle after the threat has occurred. Assuming the vehicle has  $n$  components which are critical to the completion of its mission, the damage state vector is, in fact, an  $n$ -tuple, each element of which takes on the value 0 or 1, indicating the state of the component as either functional or non-functional. If the vehicle has  $n$  critical components, then the maximum number of possible damage states is  $2^n$ . Statistical tests used to compare output from the stochastic simulation model with results of live-fire tests have, in the past, estimated the distribution of the damage states by considering the damage state vectors obtained in 1000 replications of the model. Checking to see whether or not the damage state observed in the live-fire test falls in the tail of the resulting empirical distribution function has allowed for a decision on the desired consistency question. However, the following two problems have arisen:

- 1) There are instances when the live-fire result does not match the output from any individual replication of the simulation.

- 2) There are instances when many (much greater than 5%) of the replications of the simulation output a unique damage state, thus making the tail of the empirical distribution function rather nebulous.

For both problems there has been concern that the number of replications of the simulation model may have been too small. However, to date there has been no guidance concerning the number to which it should be increased. Assuming we know the probability of kill ( $p$ ) for each of  $n$  independent components in a damage state, we have been able to determine how many times the simulation must be replicated to be  $x\%$  confident that we have seen  $y\%$  of the distribution of the damage states [1]. However, for many values of  $p$  and  $n$  and typical values of  $x$  and  $y$ , this number of replications is impracticable. For instance, suppose that there are ten critical components in a damage state vector, implying a total of 1024 possible different damage states. This might be a reasonable number of critical components for a particular compart-

ment of the vehicle, such as the engine compartment. Also, suppose that each of the ten components has a probability of kill equal to 0.5. This is probably a less reasonable assumption. Then if we consider 1000 replications of the simulation model, we would expect to see 638 different damage states representing only 62% of the total distribution. If we were considering just five critical components each with a probability of kill equal to 0.5, then we would expect to see 32 (100%) different damage states in 1000 replications of the simulation model.

As another example, suppose that there are only five critical components in a damage state vector, implying a total of 32 possible different damage states. Furthermore, suppose we would like to be sure that we have seen at least 25 of these damage states (approximately 78%). Then, if each component has a probability of kill equal to 0.5, we would need only about 50 replications of the simulation model. This number increases to about 100 if all probabilities of kill are equal to 0.7 and further increases to about 2500 if all probabilities of kill are equal to 0.9. Of course, in general, the probabilities will not be equal for any given group of critical components, but analogous results should follow. They are driven by the probability of the least likely damage state. In the first case the least likely damage state has a probability of occurrence equal to 0.03; in the second case that probability is 0.002; and in the third case that probability is 0.00001.

Thus, the results are affected by both the number of critical components in the damage state vector and the probability of the least likely damage state. With the former a large number of components implies more possible combinations for the damage state vector; with the latter a small probability of occurrence for a particular damage state implies more trials before a success is likely. These examples show that in many situations the answer to the question, "*How many replications of the simulation model are necessary?*" is "*A large number!*" Since it is often not feasible to replicate the simulation model the number of times sufficient to establish a good estimate of the distribution of damage states, we ask the following question:

Are there statistical techniques that allow us to address the problems mentioned above when the number of simulation model replications necessary to adequately describe the distribution of the damage state vectors is impracticable?

## **Possible Approaches**

There have been remedial approaches suggested for both problems. The first problem has been handled by treating the live-fire result never matched in any replication of the simula-



tion model as falling in the tail of the distribution for the purpose of testing the null hypothesis. This is done, even though a single occurrence on the 1001st replication might have pushed the question of whether or not to reject the null hypothesis into the realm of the second problem. Currently, the second problem forces us into a no-test situation, in that we are unable to apply our statistical test to examine the null hypothesis. One suggestion for overcoming this problem has been to apply bootstrapping to our 1000 outcomes from the simulation model in order to obtain an empirical distribution function with a more clearly defined tail. At the present time, we have yet to pursue this proposal.

## **Panel Suggestions**

The clinical panelists focused on the issue of independence. The consensus seemed to be that if the components are indeed dependent, then the problem can be solved only with strong prior information using a Bayesian approach. If such information is not available, then the suggestion was to group the components so that they are independent within a group, perform a separate test for each group, and combine the results of these tests to obtain an overall test statistic which may or may not allow for the rejection of the null hypothesis.. There were additional comments from both the panelists and the audience, including a suggestion to examine the literature pertaining to importance sampling for rare events.

## **Reference**

- [1] Lawrence D. Losie, "Examination of the Distribution of the Number of Component-Damage States", Army Research Laboratory Report, In Preparation.

## Formalizing the Determination of Spall Cone Angle

Barry A. Bodt  
Army Research Laboratory

In certain applications in ballistic testing it is desirable to compute a rough measure of the damaging capability of spall. This measure is dependent on two inputs. One is the cone angle, the vertex angle of an assumed right circular cone representing the path of the debris cloud behind a plate. The issue is determining this angle based on the location of the target exit hole and the impact locations of spall fragments recorded in a witness plate beyond the exit hole. Current procedure leaves outlier identification and the choice of cone angle to the experience of the vulnerability analyst. This process must be formalized to make cone angle an objectively determined quantity which can be automated easily. This clinical paper was intended to generate the discussion of possible solutions to the determination of the spall cone angle.

### 1. Introduction

An important lethal characteristic of antiarmor munitions is the ability to generate spall-fragments produced by ballistic perforation of armor. Translating spall to munition effectiveness is a recurring exercise. An inexpensive characterization of spall, used when comparing munition or armor prototypes, consists of counting the number of fragments generated and determining the angle of the spall cone, a solid right circular cone having a vertex at the point of munition exit through the armor. Not all fragments observed are necessarily contained in the representative spall cone. The choice of cone angle should be consistent with the idea that there is a sufficient density of fragments to cause components of the target within the cone to be hit with high probability. Determination of the cone angle is based on a subjective judgement as to which subset of the fragment trajectories should be included in the cone angle computation. The goal of this effort is to formalize a method for determining the cone angle which is in the spirit of the subjective estimates now made, which lends itself to automation, and which will yield more consistent results over a variety of data-structures.

### 2. Problem

In order to fully understand the question posed the complete problem is presented, beginning with the data collection and concluding with the analysis. This is followed by a brief discussion of considerations and suggested approaches. The principle interest is to receive recommendations for how one might best determine a cone angle.

Data are collected and processed with respect to the spall-cone concept. A test set-up consists of a target and a witness plate. See Figure 1. Debris consisting of both

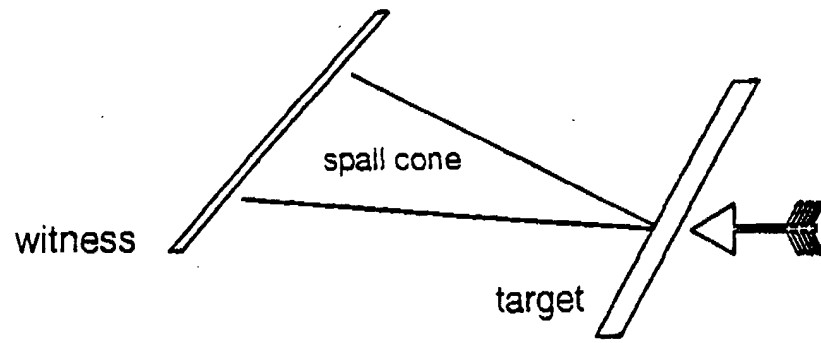


Figure 1. Test set-up.

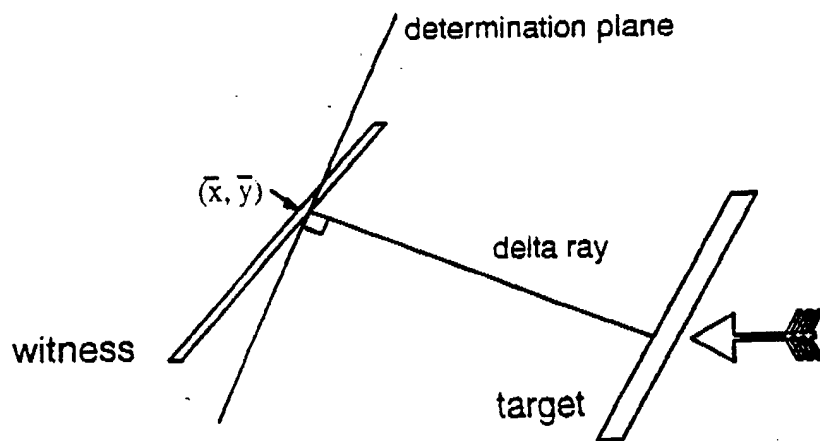


Figure 2. Determination plane.

armor fragments from the rear face of the target and penetrator fragments are generated when a penetrator perforates the target. A record of this debris for later review is created when the debris passes through a witness plate, a thin sheet of metal located behind the target. The witness plate is scanned by computer to determine the  $(x, y)$  coordinates in the witness plate plane. Through use of an iterative procedure (discussed in the appendix) points are then projected to a plane where the determination of cone angle is made. Figure 2 illustrates this "determination" plane. A normal vector (the delta ray) to the determination plane extends from the point representing the target exit hole backward to  $(x_d, y_d)$ , where the subscript d indicates that the average is taken over the point projections in the determination plane. The delta ray becomes the axis of the cone having circular base in the determination plane and vertex, the target exit hole.

After transformation, the analyst is concerned only with the point projections in the determination plane. The cone should envelop a substantive portion of the debris, consequently its circular base should include most of the  $(x, y)$  coordinates in this plane. The base center is  $(x_d, y_d)$ . Through visual inspection a central portion of the data are chosen about the center—essentially some apparent outliers may be ignored. The point,  $p$ , furthest from the center but within the central portion is identified. The angle between the delta ray and the vector extending forward from  $p$  to the target exit hole is the cone semiangle. Figures 3-4 show typical data sets with observations judged unusual noted. In each, zero represents the impact location of the largest fragment.

Once established, the cone angle, indicating how wide spread the damage might be, and the number of fragments generated, indicating the intensity of the fragment spray, are jointly used to assess effectiveness. From the perspective of munition effectiveness, the worst result—no spall—would be  $(w=0, z=0)$ , where  $w$  is the cone angle and  $z$  the number of fragments. Positive values for each show the potential for spall damage. It is further argued that if the two measures are treated as contributing equally to damage, that the distance between the origin and the point defined by the observed cone angle and number of fragments might serve as a reasonable measure of effectiveness—some scaling would be required. Comparing two munitions could then be accomplished with a univariate analysis, based on several shots, leaving the causes of significant differences to be pursued in terms of cone angle and number of fragments individually.

A second univariate measure would use a vulnerability model to predict the lethality of the munition based on cone angle and number of fragment inputs. The univariate analysis could proceed similarly.

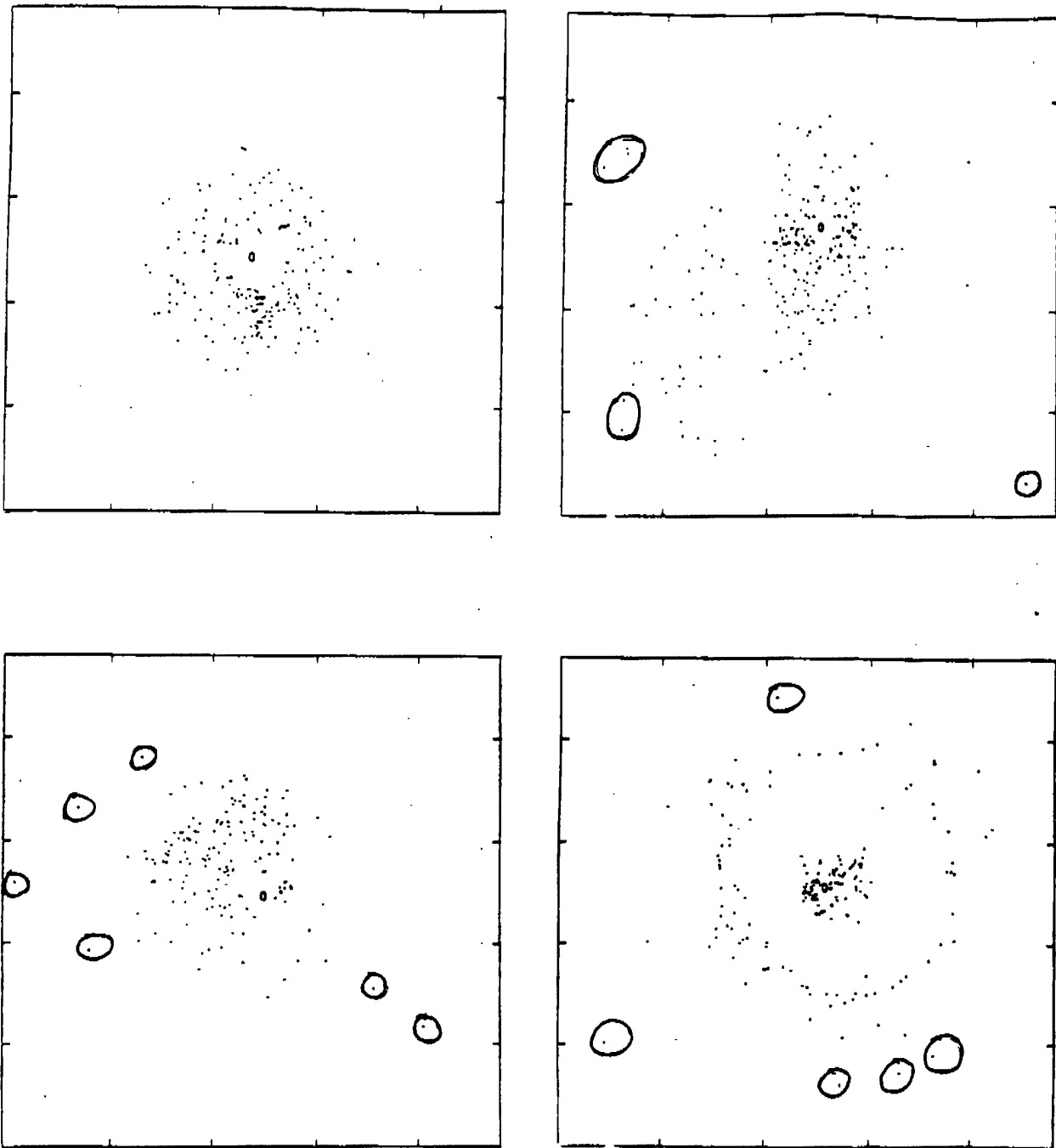


Figure 3. Four plots of representative debris with unusual points circled.

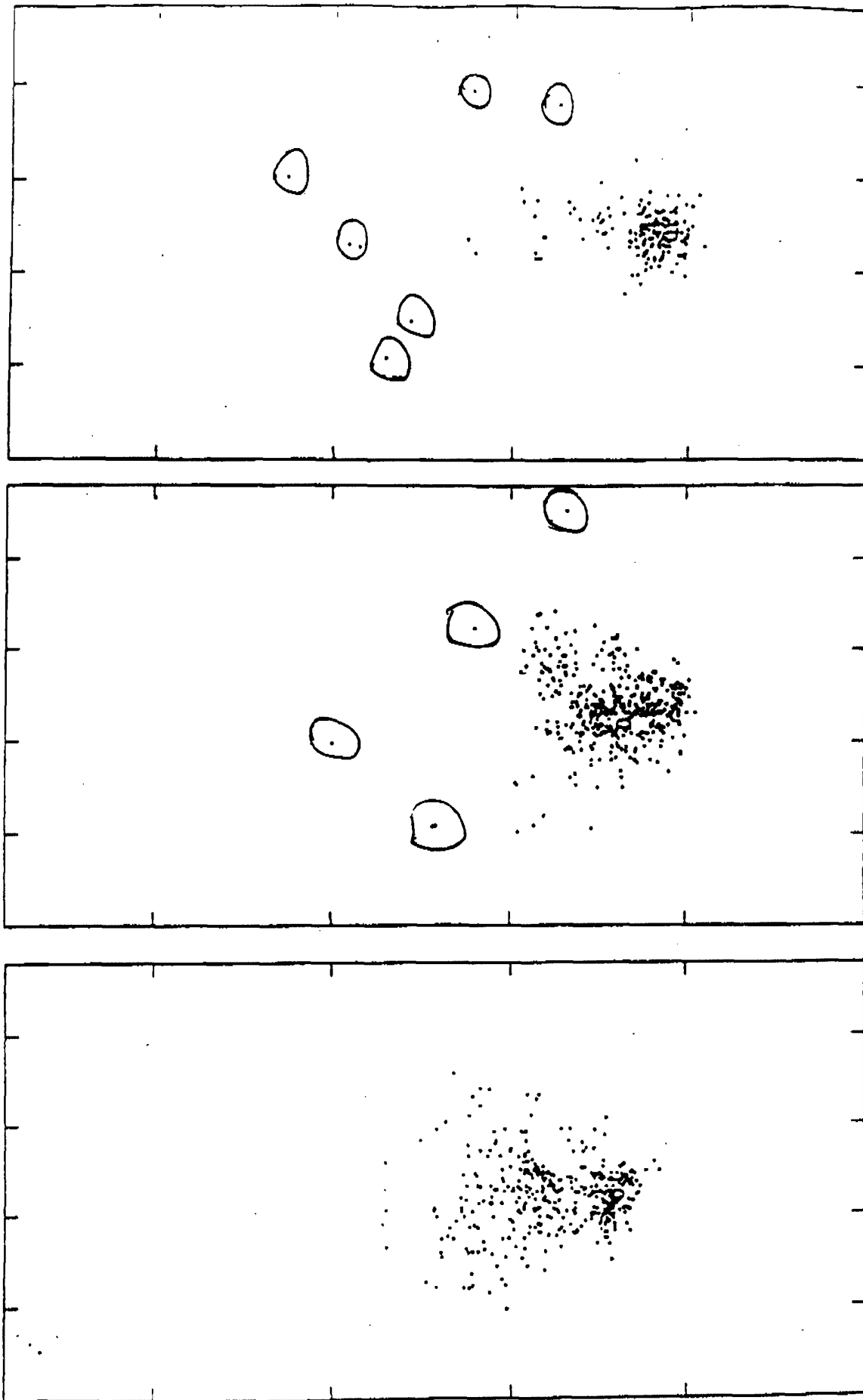


Figure 4. Three plots of representative debris with unusual points circled.

### 3. Thoughts

There are a few considerations one must be cognizant of in proposing solutions to the problem. The spall-cone concept is here to stay and is totally consistent with the way vulnerability models are constructed and interpreted. The application is intended to be a rough comparison of munition effectiveness answering How does a candidate munition's effectiveness compare against that of a baseline munition? or Where does the candidate munition fall short? A univariate measure to describe effectiveness is necessary in the sense that a ranking of performance is needed. It is insufficient to simply detect a difference between two bivariate point clouds without any means to say which is better.

Specific to the cone angle determination, we would like the angle to be accurate in some sense and to be determined consistently in a robust fashion. The subjective manner in which the cone is now established is troublesome from the viewpoint of consistency; although, relying on the judgement of experts on the damaging capability of spall, is in a sense accurate. Some ideas proposed follow.

One suggestion is to not look at the bivariate data to determine the cone angle. Rather, use each point in the determination plane to form a possible cone semiangle. An empirical distribution of cone semiangles could now be studied for the presence of outliers.

Another suggestion, if the bivariate data were normal is to consider the distance from the center, distributed as a Rayleigh. Perhaps the cone semiangle could be based on some quantile of that distribution. Some believe that a bivariate Weibull or normal model would describe the data well. Perhaps other covering-circle approaches would be appropriate. A caution with this suggestion and the preceding might be that points appearing to be outliers in a scatterplot might not show up as unusual in terms of a single univariate measure, perhaps resulting in an artificially large cone angle.

Responding to that concern, another suggestion is to divide the bivariate cloud into angular segments, perhaps looking for the maximum distance that maintains a reasonable intensity of points as you move away from the center along each of the segments.

### 4. Panel Discussion

The panel discussion was most worthwhile. The session chairman, Terry Cronin, and the discussants Nozer Singpurwalla, James Hodges, David Scott, and Donald Gaver each had suggestions. Their offerings are not detailed here but briefly one used a Bayesian approach involving a bivariate t-distribution to model the impact locations on the witness plate. Another suggestion was to use nonparametric density

estimation to establish an area of damage falling within some quantile on the nonparametric density. It was also demonstrated how one could cover the data using concentric circles about the center with increasing radii until a stopping rule was satisfied. Independently, two recommended that outliers could be sequentially removed to conform with an angle requirement for vertices of an outer convex hull formed about the data. Each of the approaches mentioned has been summarized and included along with other proposals for review by a working group formed to study this issue.

## Appendix

For the purposes of this discussion, the panel can view the witness plate plane as being the plane in which the circular base of the cone resides. However, for completeness the original plane defined by the witness plate is not usually the plane in which the cone angle determination is made. A normal vector to the witness plate plane is established, passing through the point marking the exit of the penetrator. This vector is termed the delta ray. Conceptually, the delta ray represents the central path of the debris. Then the average of the  $(x, y)$  coordinates is taken on the witness plate plane. If this average differs substantially from the point where the delta ray intersects the plane, several steps are taken. First, a new delta ray is formed extending from the target plate exit hole to  $(x, y)$ . Second, a new plane is defined such that the new delta ray is normal to it and  $(x, y)$  resides on it. Third, points from the original witness plate plane are projected to the new plane by determining the point of intersection between the new plane and the vector extending from the target exit hole to the point in the witness plate plane. A second iterate of this process begins with a computation of  $(x, y)$  in the new plane. Several iterations may be required to achieve a delta ray which is normal to the determination plane, passes through  $(x, y)$  in that plane, and passes through the point marking the target exit hole.



Revised 20th February 1992

# **Models for Assessing the Reliability of Computer Software**

**Nozer D. Singpurwalla and Simon P. Wilson**  
**The George Washington University, Washington, D.C. 20052**

**GWU/IRRA/Serial TR-91/13**  
**December 1991**

**Research Supported by**

**Contract N00014-85-K-202**  
**Office of Naval Research**

**Grant DAAL03-87-K-0056**  
**The Army Research Office**

**and**

**Grant AFOSR-F49620-92-J-0030**  
**The Air Force Office of Scientific Research**

# Models for Assessing the Reliability of Computer Software

Nozer D. Singpurwalla and Simon P. Wilson

The George Washington University, Washington, D.C. 20052

## Abstract

A formal approach for evaluating the reliability of computer software is through probabilistic models and statistical methods. This paper is an expository overview of the literature on the former. The various probability models for software failure can be classified into two groups; the merits of these groups are discussed and an example of their use in decision problems is given in some detail. The direction of current and future research is contemplated.

# 1. Introduction.

Having been developed over the last 20 years, software reliability is a relatively new area of research for the statistics and the computer science communities. It arose because of interest in trying to predict the reliability of software, particularly when its failure could be catastrophic. Obviously the software that controls an aircraft carrier, a nuclear power station, a submarine or a life-support machine needs to be very reliable, and statistical techniques will aid the computer scientist in deciding if such software has sufficient reliability. The subject is also of commercial importance, as for example when decisions have to be made concerning the release of software into the marketplace.

All software is subject to failure, due to the inevitable presence of errors (or bugs) in the code, so the first aim of the subject has been to develop models that describe software failure. There are various methods of specifying such failure models, and Section 2 discusses these in some detail. It is fair to say that this model derivation has been the focus of research so far. Once a failure model has been specified then it can be applied to problems such as the optimal time to debug software or deciding whether software is ready for release. These applications have received less attention in the literature but are becoming more prevelant. We will mention here that there is another approach to software reliability that differs considerably from the statistical ideas presented here. This approach attempts to prove the reliability, or correctness, of software by formal means of proof, just as one would prove a mathematical theorem. This is an exercise in logic, albeit a rather complex one. It works well on small programs, for example on a program that computes the factorial function, but becomes a forbidding task for even moderately complex pieces of code. Nevertheless, the idea that software can be proved correct is appealing. The approach is not discussed further.

This paper is divided into 5 further sections. Section 2 categorizes the different strategies that have been used to model software failure. Section 3 reviews the historical development of the subject by describing some of the more commonly used models, and Section 4 shows that many of these models can be unified if one adopts a Bayesian position. Section 5 looks at applications of the material developed in Section 3, and Section 6 concludes with a look at the current and future direction of the subject. We assume that the reader has some familiarity with some basic reliability and probability concepts; in particular it is important that he or she has knowledge of some common probability distributions, statistical inference and decision making, Poisson processes and the concept of a failure rate.

## 2. Model Categorization

All statistical software reliability models are probabilistic in nature. They attempt to specify the probability of software failure in some manner. In looking through the literature, one observes that the models developed so far can be broadly classified into two categories

**Type I:** Those which propose a probability model for *times between successive failure* of the software, and

**Type II:** Those which propose a probability model for the *number of failures* up to a certain time.

Time is often taken to be CPU time, or the amount of time that the software is actually running, as opposed to real time. In theory, specification via one of these two methods enables one to specify the other. So a model that specifies time between failure will also be able to tell you the number of failures in a given time, and vice versa. In practice, this may not be straightforward.

The first of these categories, modeling time between failure, is most commonly accomplished via a specification of the *failure rate* of the software as it is running. When this is the case the model is to be of **Type I-1**. The failure rate for the  $i$ -th time between failure is given, for  $i=1, 2, 3, \dots$  and a probability model results. One distinctive feature of software is that its failure rate may decrease with time, as more bugs are discovered and corrected. This contrasts with most mechanical systems which will age over time and so have an increasing failure rate. An attempt to debug software may introduce more bugs into it, thus tending to increase the failure rate, so the decreasing failure rate assumption is somewhat idealized. However, most of the models of this type that are reviewed here have a decreasing failure rate.

Another way to model time between failure is to define a stochastic relationship between successive failure times. Models that are specified by this method are known as **Type I-2**, and have the advantage over Type I-1 in that they model the times between failure directly, and not via the abstract concept of a failure rate. For example, let  $T_1, T_2, \dots, T_i, \dots$  be the length of times between successive failure of the software. As a simple case, one could declare that  $T_{i+1} = \rho T_i + \epsilon_i$ , where  $\rho \geq 0$  is a constant and  $\epsilon_i$  is an error term (typically some random variable with mean 0). Then  $\rho < 1$  would indicate decreasing times between failure (software reliability expected to become worse),  $\rho = 1$  would indicate no expected change in software reliability whilst  $\rho > 1$  indicates increasing times between failure

(software reliability expected to improve). Those familiar with time series will recognize the relationship in this example as an auto-regressive process of order 1; in general, one would say  $T_{i+1} = f(T_1, T_2, \dots, T_i) + \epsilon_i$  for some function  $f$ .

The second of these categories, modeling the number of failures, uses a point process to count the failures. Let  $M(t)$  be the number of failures of the software that are observed during time  $[0, t]$ .  $M(t)$  is modeled by a *Poisson process*, which is a stochastic process with the following properties:

- i)  $M(0) = 0$  and if  $s < t$  then  $M(s) \leq M(t)$ .  $M(t)$  takes values in  $\{0, 1, 2, \dots\}$
- ii) The number of failures that occur in disjoint time intervals are independent. So, for example, the number of failures in the first 5 hours of use has no effect on the number of failures in the next 5 hours.
- iii) The number of failures to time  $t$  is a Poisson random variable with mean  $\mu(t)$ , for some non-decreasing function  $\mu(t)$ ; that is to say:

$$P(M(t)=n) = \frac{(\mu(t))^n}{n!} e^{-\mu(t)} \quad n=0, 1, 2, \dots$$

The different models of this type have a different function  $\mu(t)$ , which is called the mean value function. The mean number of failures at time  $t$  is indeed  $\mu(t)$ , as is the variance. The Poisson process is chosen because in many ways it is the simplest point process yet it is flexible and has many useful properties that can be exploited. This second approach has become increasingly popular in recent years.  $M(t)$  can also be specified by its intensity function  $\lambda(t)$ , which is the derivative of  $\mu(t)$  with respect to  $t$ ; either of these functions completely specify a particular Poisson process. One disadvantage of this approach is that it implies that there are conceptually an infinite number of bugs in the program, which is obviously impossible for code of a finite length. Another disadvantage is more subtle; the model implies a positive correlation between the number of failures in adjoining time intervals, a situation which is not true since again the total number of bugs has to be finite.

Figure 1 is a flow-chart showing the above categorization of the statistical models.

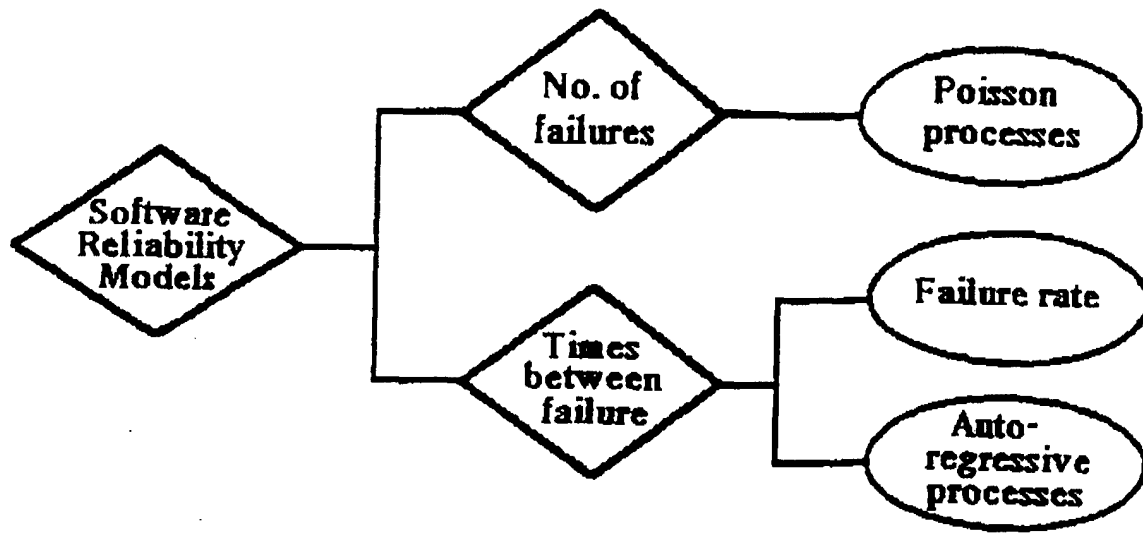


Figure 1. Categorization of Software Reliability Models.

### 3. Review of Some Software Reliability Models

This section introduces some of the more well known probability models for software reliability. There are examples of models from each of the two main categories that were discussed in the previous section. Since the main purpose of the review is to describe the ideas and assumptions behind the models, technical details will be kept to a minimum in most cases. Those interested in the details of a particular model are advised to reference the papers where they were originally presented.

Some common notation will be assumed throughout this section and is given below :

- i)  $T_i$  = i-th time between failure of the software [i.e. time between (i-1)th and i-th failure].
- ii)  $r_{T_i}(t)$  = failure rate for  $T_i$ , the i-th time between failure, at time t.
- iii)  $M(t)$  = number of failures of the software in the time interval  $[0, t)$  (a Poisson process).
- iv)  $\lambda(t)$  = intensity function of  $M(t)$ .
- v)  $\mu(t)$  = expected number of failures of software in time  $[0, t)$ .

$$= \int_0^t \lambda(s) ds, \text{ since } M(t) \text{ is a Poisson process.}$$

10 models are presented. Model numbers 1 to 7 are of Type I-1, models 8 and 9 are of Type II and model 10 is of type I-2. A common problem to all the models is the lack of data on which to test their validity; data on software reliability is commercially sensitive and so statisticians in academia have very little information on how software in the marketplace actually performs. For this reason it is important that the assumptions made in deriving these models are carefully thought about.

#### 1. The model of Jelinski & Moranda (1972).

This was the very first software reliability model that was proposed, and has formed the basis for many models developed after. It is a Type I-1 model; it models times between failure by considering their failure rates. Jelinski and Moranda reasoned as follows. Suppose that the total number of bugs in the program is  $N$ , and suppose that each time the software fails, one bug is corrected. The failure rate of the i-th time between failure,  $T_i$ , is then assumed a constant proportional to  $N-i+1$ , which is the number of bugs remaining in the program. In other words

$$r_{T_i}(t | N, \Lambda) = \Lambda (N-i+1), \quad i=1, 2, 3, \dots, t \geq 0, \quad \text{for some constant } \Lambda.$$

There are some criticisms that one could make of the model. It assumes that each error contributes the same amount  $\Lambda$  to the failure rate, whereas in reality different bugs will have different effects. It also assumes that every time a fix is made, no new bugs are introduced: note [see Figure 2(i)] that the successive failure rates are indeed decreasing. A model like this is sometimes referred to as a "de-entrophication model", because the process of removing bugs from software is akin to the removal of pollutants in rivers and lakes.

## 2. Bayesian Reliability Growth Model (Littlewood & Verall (1973)).

Like the Jelinski & Moranda model, the model proposed by Littlewood and Verall looked at times between failure of the software. However, they did not develop the model by characterizing the failure rate; rather they stated that the model should *not* be based on fault content (as Jelinski & Moranda had assumed) and then declared that  $T_i$  has an exponential distribution with scale  $\Lambda_i$ , and that  $\Lambda_i$  itself has a gamma distribution with shape  $\alpha$  and scale  $\Psi(i)$ , for some function  $\Psi$ . Despite this it is still considered to be a Type I-1 model.

Specifically :

$$f_{T_i}(t | \Lambda_i) = \Lambda_i e^{-\Lambda_i t} \quad t \geq 0$$

$$\Pi_{\Lambda_i}(\lambda | \alpha, \Psi(i)) = \frac{(\Psi(i))^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\Psi(i)\lambda} \quad \lambda \geq 0$$

$\Psi(i)$  was supposed to describe the quality of the programmer and the programming task. As an example, they chose  $\Psi(i) = \beta_0 + \beta_1 i$ . One can show that this makes the failure rate of each  $T_i$  decreasing in  $t$  and that each time a bug is discovered and fixed there is a downward jump in the successive failure rates; see Figure 2(ii). In fact

$$r_{T_i}(t | \alpha, \beta_0, \beta_1) = \frac{\alpha}{\beta_0 + \beta_1 i + t}, \quad \text{for } t \geq 0.$$

If  $\beta_1 > 1$  then the jumps in the failure rate decrease in  $i$ , if  $\beta_1 < 1$  they increase whilst if  $\beta_1 = 1$  they remain a constant. So if  $\beta_1$  differs from 1 then the fixing of each bug is making a different contribution to the reduction in the failure rate of the software, an apparent advantage over the model by Jelinski &



Moranda. This model has received quite a lot of attention and has been the subject of various modifications: see models 6 and 7 later in this section.

### 3. The De-eutrophication model of Moranda (1975).

Another (de-eutrophication) model of Moranda (1975) attempted to answer some of the criticisms of the Jelinski & Moranda model, in particular the criticism concerning the equal effect that each bug in the code has on the failure rate. He hypothesized that the fixing of bugs that cause early failures in the system reduces the failure rate more than the fixing of bugs that occur later, because these early bugs are more likely to be the bigger ones. With this in mind, he proposed that the failure rate should remain constant for each  $T_i$ , but that it should be made to decrease geometrically in  $i$  after each failure i.e. for constants  $D$  and  $k$

$$r_{T_i}(t | D, k) = D k^{i-1} \quad t \geq 0, D > 0 \text{ and } 0 < k < 1.$$

Compared to the Jelinski & Moranda model, where the drop in failure rate after each failure was always  $\Lambda$ , the drop in failure rate here after the  $i$ -th failure is  $D k^{i-1}(1-k)$  see Figure 2(iii). The assumption of a perfect fix, with no introduction of new bugs during the fix, is retained.

### 4. Imperfect Debugging Model (Goel & Okumoto (1978)).

This model is another generalization of the Jelinski & Moranda model which attempts to address the criticism that a perfect fix of a bug does not always occur. Goel & Okumoto's *Imperfect Debugging Model* is like the Jelinski & Moranda model, but assumes that there is a probability  $p$ ,  $0 \leq p \leq 1$ , of fixing a bug when it is encountered. This means that after  $i$  faults have been found, we expect  $i \times p$  faults to have been corrected, instead of  $i$ . Thus the failure rate of  $T_i$  is

$$r_{T_i}(t | N, \Lambda, p) = \Lambda (N - p(i-1))$$

When  $p=1$  we get the Jelinski & Moranda model.

##### 5. A model by Schick & Wolverton (1978).

This is yet another Type I model, and this time the failure rate is assumed proportional to the number of bugs remaining in the system and the time elapsed since the last failure. Thus

$$r_{T_i}(t | N, \Lambda) = \Lambda (N-i+1)t, \quad t \geq 0$$

This model differs from models 1-4 in that the failure rate does not decrease monotonically. Immediately after the  $i$ -th failure, the failure rate drops to 0, and then increases linearly with slope  $(N-i)$  until the  $(i+1)$ th failure: see Figure 2(iv).

##### 6. Bayesian Differential Debugging Model (Littlewood (1980)).

This model can be considered as an elaboration of model 2 proposed by Littlewood & Verall. Recall that in model 2 it was assumed that  $\Lambda_i$ , the failure rate of the  $i$ -th time between failures, was declared to have a gamma distribution. In this new model Littlewood supposed that there were  $N$  bugs in the system (a return to the bug counting phenomenon), and then proposed that  $\Lambda_i$  be specified as a function of the remaining bugs. In particular, he stated  $\Lambda_i = \phi_1 + \phi_2 + \dots + \phi_{N-i}$ , where  $\phi_i$  were independent and identically distributed gamma random variables with shape  $\alpha$  and scale  $\beta$ . This implied that  $\Lambda_i$  would have a gamma distribution with shape  $\alpha(N-i)$  and scale  $\beta$ . In other respects its assumptions are identical to the original Littlewood/Verall model.

##### 7. Bayes Empirical Bayes or Hierarchical Model (Mazzuchi & Soyer (1988)).

In 1988 Mazzuchi and Soyer proposed a *Bayes Empirical Bayes* or *Hierarchical* extension to the Littlewood & Verall model (model 2). As with the original model, they assumed  $T_i$  to be exponentially distributed with scale  $\Lambda_i$ . Then they proposed two ideas for describing  $\Lambda_i$ , here called model A and model B.

###### Model A :

Still assume that  $\Lambda_i$  is described by a gamma distribution, but with parameters  $\alpha$  and  $\beta$ . Now assume that  $\alpha$  and  $\beta$  are independent and that they themselves are described by probability distributions;  $\alpha$  by a uniform and  $\beta$  by another gamma. In other words :

$$\Pi_{\Lambda_i}(\lambda | \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}, \quad \lambda \geq 0$$

$$\pi(\alpha | \nu) = \frac{1}{\nu}, \quad 0 \leq \alpha \leq \nu$$

$$\pi(\beta | a, b) = \frac{b^a}{\Gamma(a)} \beta^{a-1} e^{-b\beta}, \quad \beta \geq 0, a > 0, b > 0.$$

Model B:

Assume that  $\Lambda_i$  is described exactly as in Littlewood and Verall i.e.

$$\Pi_{\Lambda_i}(\lambda | \alpha, \Psi(i)) = \frac{(\Psi(i))^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\Psi(i)\lambda}, \quad \lambda \geq 0$$

and that  $\Psi(i) = \beta_0 + \beta_1 i$ , except now place probability distributions on  $\alpha$ ,  $\beta_0$  and  $\beta_1$  as follows:

$$\pi(\alpha | \omega) = \frac{1}{\omega}, \quad 0 \leq \alpha \leq \omega$$

$$\pi(\beta_0 | a, b, \beta_1) = \frac{b^a}{\Gamma(a)} (\beta_0 + \beta_1)^{a-1} e^{-b(\beta_0 + \beta_1)}, \quad \beta_0 \geq -\beta_1, a > 0, b > 0$$

$$\pi(\beta_1 | c, d) = \frac{d^c}{\Gamma(c)} \beta_1^{c-1} e^{-d\beta_1}, \quad \beta_1 \geq 0, c > 0, d > 0.$$

So  $\alpha$  is described by a uniform distribution,  $\beta_0$  by a shifted gamma and  $\beta_1$  by another gamma, and there is dependence between  $\beta_0$  and  $\beta_1$ . By assuming  $\beta_1$  to be degenerate at 0, model A is obtained from model B. The authors were able to find an approximation to the expectation of  $T_{n+1}$  given that  $T_1=t_1, T_2=t_2, \dots, T_n=t_n$ , and so use their model to predict future reliability of the software in light of the previous failure times.

## 8. Time-dependent Error Detection Model (Goel & Okamoto (1979)).

This is the first Type II model that we will consider. It assumes that  $M(t)$ , the number of failures of the software in time  $[0, t]$ , is described by a Poisson process with intensity function given by

$$\lambda(t) = ab e^{-bt}$$

where  $a$  is the total expected number of bugs in the system and  $b$  is the fault detection rate; see Figure 2(v). Thus the expected number of failures to time  $t$  is :

$$\mu(t) = \int_0^t ab e^{-bs} ds = a (1 - e^{-bt}).$$

The function  $\mu(t)$  completely specifies a particular Poisson process, and the distribution of  $M(t)$  is given by the well known formula

$$P(M(t)=n) = \frac{(\mu(t))^n}{n!} e^{-\mu(t)}, \quad n=0,1,2,\dots$$

Experience has shown that often the rate of faults in software increases initially before eventually decreasing, and so in Goel (1983) the model was modified to account for this by letting

$$\lambda(t) = abc t^{c-1} e^{-bt^c}$$

where  $a$  is still the total number of bugs and  $b$  and  $c$  describe the quality of testing.

#### 9. Logarithmic Poisson Execution Time Model (Musa and Okumoto (1984)).

The *Logarithmic Poisson Execution Time Model* of Musa and Okumoto is one of the more popular software failure models of recent years. It is a type II model, but the model is not derived by directly assuming some intensity function  $\lambda(t)$ , as was the case with model 8 of Goel & Okumoto. Here  $\lambda(t)$  is expressed in terms of  $\mu(t)$ , the expected number of failures in time  $[0,t)$ , via the relationship

$$\lambda(t) = \lambda_0 e^{-\theta\mu(t)}.$$

Put simply, this relationship encapsulates the belief that the intensity (or rate) of failures at time  $t$  decreases exponentially with the number of failures experienced, and so bugs fixed, up to time  $t$ . The fixing of earlier failures will reduce  $\lambda(t)$  more than the fixing of later ones. Since we are modeling the number of failures by a Poisson process, then we have another relationship between  $\lambda(t)$  and  $\mu(t)$ , namely

$$\mu(t) = \int_0^t \lambda(s) ds.$$

Using these two relationships between  $\lambda(t)$  and  $\mu(t)$ , there is a unique solution for the two functions:

$$\lambda(t) = \frac{\lambda_0}{\lambda_0\theta t + 1} \quad ; \quad \mu(t) = \frac{1}{\theta} \ln(\lambda_0\theta t + 1).$$

Figure 2 (vi) shows a plot of  $\lambda(t)$  versus  $t$ : it is similar to the plot of figure 2 (v) except that the tail is thicker.

It now follows from the above that by using  $P(M(t)=n) = (\mu(t))^n e^{-\mu(t)} / n!$  we can say

$$P(M(t)=n) = \frac{(\ln(\lambda_0 \theta t + 1))^n}{\theta^n (\lambda_0 \theta t + 1)^{1/\theta} \times n!}, \quad n=0,1,2,\dots$$

As a final remark, we mention that in their paper the authors go into some detail on estimation of  $\lambda_0$  and  $\theta$  by maximum likelihood methods; however, one of the likelihoods appears to be incorrect.

#### 10. Random Coefficient Autoregressive process model (Singpurwalla & Soyer (1985)).

This is a Type I-2 model, that is one that does not consider the failure rate of times between failure. Instead it assumes that there is some pattern between successive failure times and that this pattern can be described by a functional relationship between them. The authors declare this relationship to be of the form

$$T_i = T_{i-1}^{\theta_i}, \quad i=1,2,3,\dots$$

where  $T_0$  is the time to the first failure and  $\theta_i$  is some unknown coefficient. If all the  $\theta_i$ 's are bigger than 1 then we expect successive lifelengths to increase, and if all the  $\theta_i$ 's are smaller than 1 we expect successive lifelengths to decrease.

Uncertainty in the above relationship is expressed via an error term  $\delta_i$ , so that

$$T_i = \delta_i T_{i-1}^{\theta_i}.$$

The authors then make the following assumptions, which greatly facilitate the analysis of this model. They assume the  $T_i$ 's to be lognormally distributed, that is to say that  $\log T_i$ 's have a normal distribution, and that they are all scaled so that  $T_i \geq 1$ . The  $\delta_i$ 's are also assumed to be lognormal, with median 1 and variance  $\sigma_1^2$  (the conventional notation is  $\Lambda(1, \sigma_1^2)$ ). Then by taking logs on the relationship above they obtain

$$\log T_i = \theta_i \log T_{i-1} + \log \delta_i$$

$$= \theta_i \log T_{i-1} + \epsilon_i, \text{ say.}$$

Since the  $T_i$ 's and the  $\delta_i$ 's are lognormal so the  $\log T_i$ 's and the  $\epsilon_i$ 's ( $= \log \delta_i$ 's) will be normally distributed, and in particular  $\epsilon_i$  has mean 0 and some variance  $\sigma^2$  (the conventional notation is  $N(0, \sigma^2)$ ). The log-lifetimes therefore form what is known as an *autoregressive process of order 1 with random coefficients*  $\theta_i$ . There is an extensive literature on such processes which can now be used on this model.

All that remains to do is to specify  $\theta_i$ , and the authors consider several alternative models. For example, one could make  $\theta_i$  itself an autoregressive process :

$$\theta_i = \alpha \theta_{i-1} + \omega_i \quad \text{where } \omega_i \text{ is } N(0, W_i) \text{ with } W_i \text{ known.}$$

When  $\alpha$  is known, the expressions for  $\log T_i$  and  $\theta_i$  together form a *Kalman filter model*, on which there is also an extensive literature. When  $\alpha$  is not known the solution is via an *adaptive Kalman filter* algorithm for which the above authors propose an approach. As an alternative to the above, one could place a two stage distribution on  $\theta_i$ , and the authors considered the idea of  $\theta_i$  being  $N(\lambda, \sigma_2^2)$ , with  $\lambda$  also a normal random variable having mean  $m_0$  and variance  $s_0^2$ . In this latter case one can employ standard hierarchical Bayesian inference techniques to predict future reliability in the light of previous failure data.

Figure 2 shows the various failure rates for models 1, 2, 3 and 5, and the intensity function for models 8 and 9.

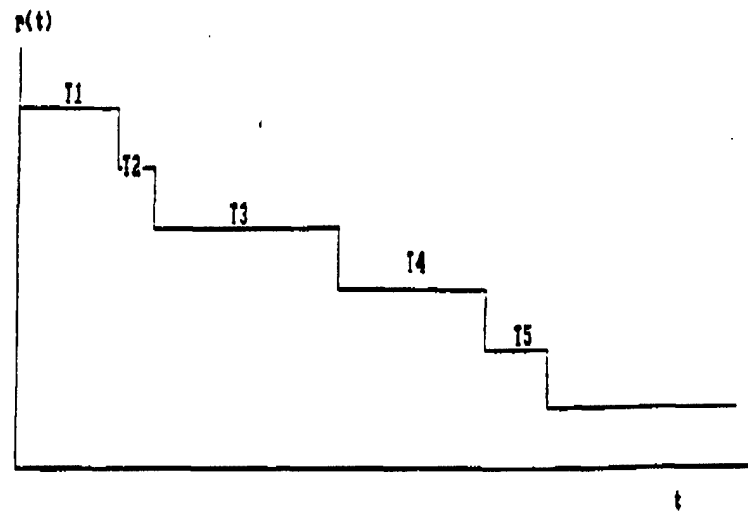


Figure 2 (i) The failure rate of the model of Jelinski and Moranda

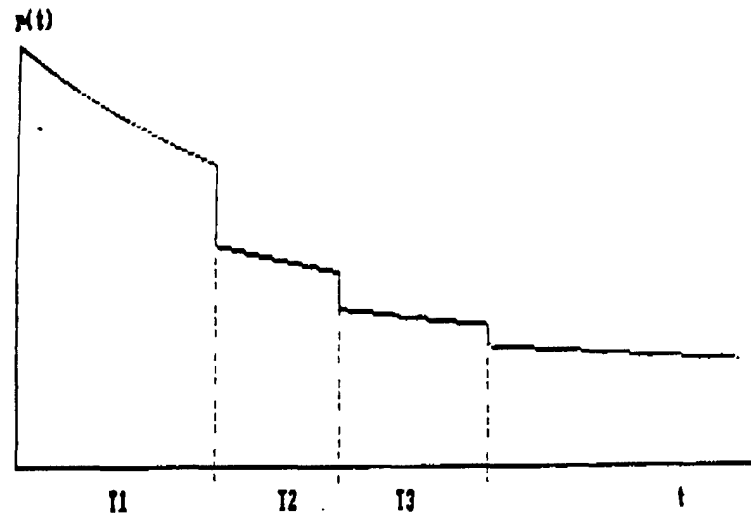


Figure 2 (ii) The failure rate of the model of Littlewood and Verall

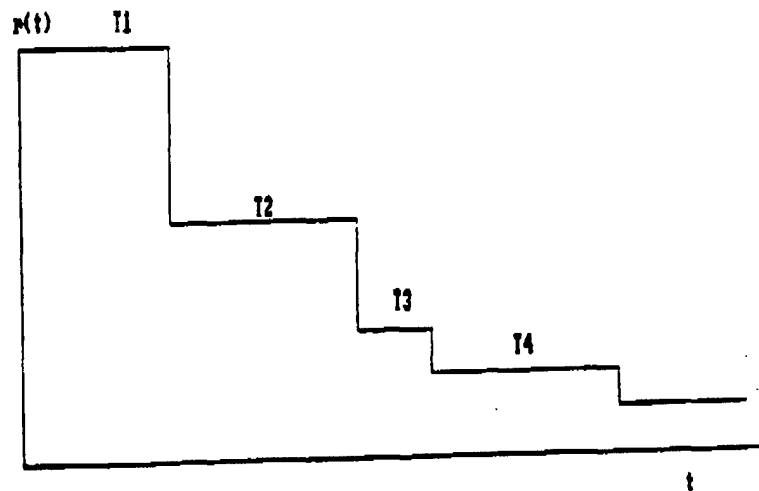


Figure 2 (iii) The failure rate of the model of Moranda

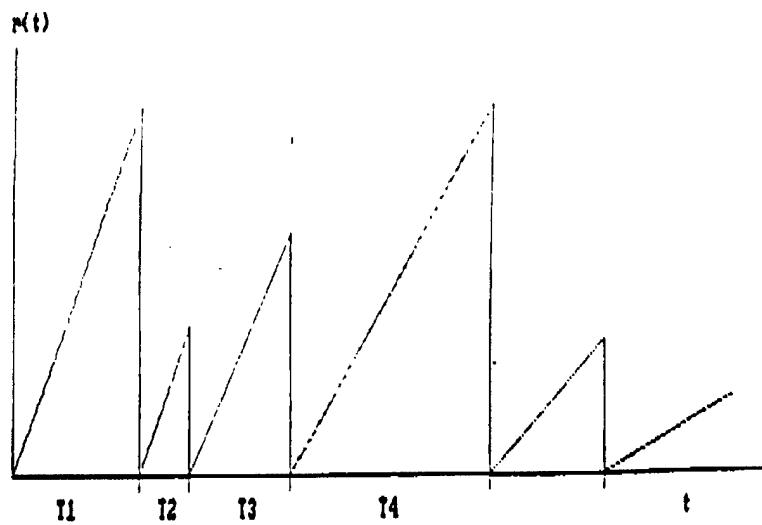


Figure 2 (iv) The failure rate of the model of Schick and Wolverton

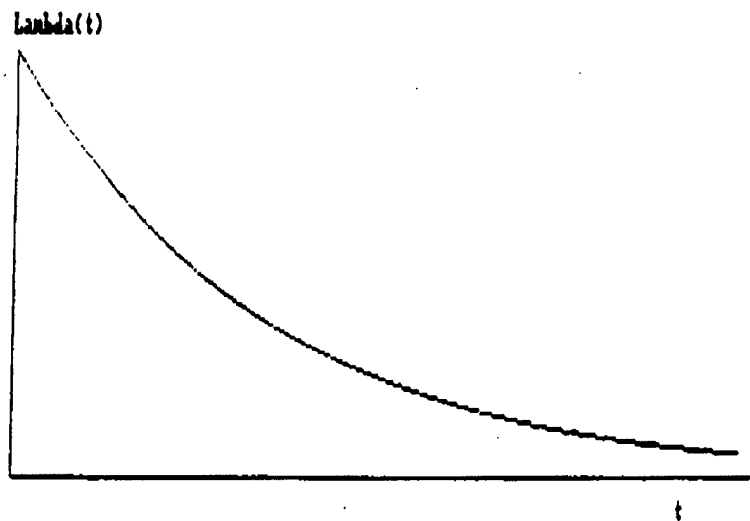


Figure 2 (v) The intensity function for the model of Goel and Okumoto

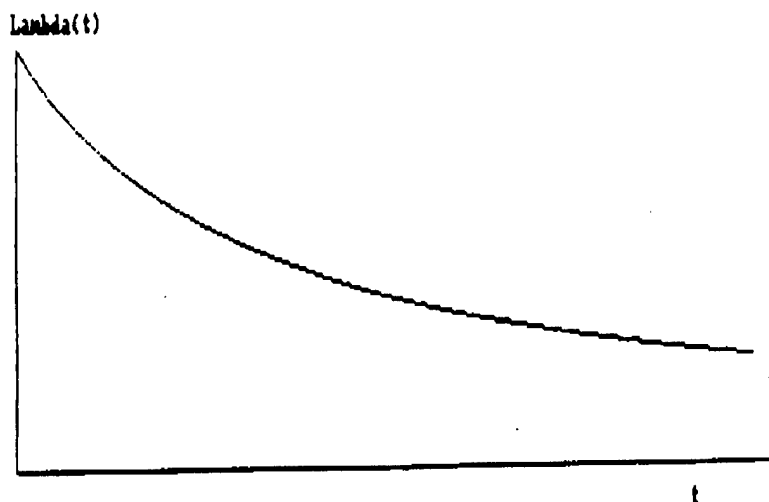


Figure 2 (vi) The intensity function for the model of Musa and Okumoto



## 4. Model Unification.

By adopting a Bayesian approach, it turns out that one can unify models 1, 2 and 8 - the models by Jelinski & Moranda, Littlewood & Verrall and Goel & Okomuto respectively - under a general framework. Observe that this also provides a link between the two types of models, since models 1 and 2 are of type I whilst model 8 is of type II.

We begin by recalling the first model, that by Jelinski & Moranda. Each  $T_i$  is assumed to have a constant failure rate  $\Lambda(N-i+1)$ . It is well known that this implies each  $T_i$  must therefore be exponentially distributed with mean  $(\Lambda(N-i+1))^{-1}$ . Now assume that  $\Lambda$  and/or  $N$  is unknown; in true Bayesian fashion prior distributions are placed upon them.

To obtain model 8 by Goel & Okomuto, we let  $\Lambda$  be degenerate at  $\lambda$  and  $N$  have a Poisson distribution with mean  $\theta$ . One can calculate  $M(t)$  using the  $T_i$ 's as defined by Jelinski & Moranda, and then by averaging out over  $N$  one finds that  $M(t)$  is indeed a Poisson process with mean :

$$\mu(t) = \theta (1 - e^{-\lambda t})$$

which is the form of  $\mu(t)$  for Goel & Okomuto's model.

One can also obtain model 2 by assuming  $N$  to be degenerate and  $\Lambda$  to have a gamma distribution. The derivations which lead to the above are complex; readers are referred to Langberg and Singpurwalla (1985) for the details.

## 5. An application : optimal testing of software.

The failure models that have been reviewed in the preceding sections can be used for more than inference or the prediction of software failure. They can also be applied in the framework of decision theory to solve decision problems. An important example of such a problem is the optimal time to test software before releasing it. This involves the balancing of the costs of testing and the risk of software obsolescence with the cost of in-service failure, should a bug not be corrected during the testing period. The following is taken from Singpurwalla (1991), in which a strongly Bayesian approach is taken.

To implement a decision theoretic procedure requires two key ingredients. The first is a probability model, and here we take a generalization of the Jelinski & Moranda model. The second is a consideration of the costs and benefits, or *utilities*, associated with a particular decision i.e the costs of testing, the benefits and costs of fixing a bug etc. Decision theory states that the optimal decision (in this case time of test) is that which *maximizes expected utility*.

If the software is to be tested for some time, say  $T$  units, and then released the problem is to find a  $T$  that maximizes expected utility. This is called *single stage testing*. There is a more complex, yet realistic, scenario called *two stage testing*. Here the software is tested for  $T$  units of time, and then depending on how many failures  $M(T)$  were observed during that test, a decision is made on whether to continue testing for a further  $T^*$  units. The problem here is to find the optimal  $T$  and  $T^*$ , with  $T^*$  to be determined before  $M(T)$  is observed. Finally there is a third testing scenario, namely *sequential testing*. Here  $T^*$  is determined after  $M(T)$  is observed; this procedure can continue for several stages, with  $T^{**}$  being determined after  $M(T^*)$  is observed and so on. Here we consider the case of single stage testing. Figure 3 is a graph of the decision process associated with single stage testing.

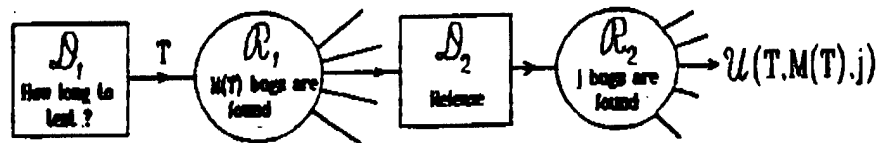


Figure 3 Decision process for single-stage testing

The model chosen in this paper is an extension to Jelinski & Moranda's model. We have

$$f_{T_i}(t | N, \Lambda) = \Lambda(N-i+1) e^{-\Lambda(N-i+1)t} \quad t \geq 0$$

In the previous section we placed prior distributions on one of  $N$  or  $\Lambda$ . Now we place priors on both the parameters, and say That  $N$  has a Poisson distribution with mean  $\theta$ ,  $\Lambda$  has a gamma distribution with scale  $\mu$  and shape  $\alpha$  and that  $N$  and  $\Lambda$  are independent.

We now turn to the choice of utility function. The following assumptions are made :

- i) The utility of a program that encounters  $j$  bugs during its operation is  $a_1 + a_2 e^{-a_3 j}$ .
- ii) The cost of fixing a bug is some constant  $C_1$ .
- iii) Let  $f(T)$  be the cost of testing and lost opportunity to time  $t$ ; here we say  $f(T) = dT^a$

Note from i) that the utility of a bug-free program is  $a_1 + a_2$ , and the utility of a program with a very large number of bugs is near  $a_1$ , so that typically  $a_1$  is a large negative number (because there is a great loss associated with software that is constantly failing in the marketplace) and  $a_2 > 0$ . Combining these assumptions gives us the utility of a program that is tested for  $T$  units of time, during which  $M(T)$  bugs are found and corrected, and then released where  $j$  bugs are encountered by the customer as

$$U(T, M(T), j) = e^{-bT} \times \{a_1 + a_2 e^{-a_3 j} - C_1 M(T) - dT^a\}$$

where  $e^{-bT}$  is some devaluating factor.

Now the two parts of the decision process - the probability model and the utility function - are brought together. We wish to find the time  $T$  that maximizes expected utility. In other words find  $\hat{T}$  such that  $E(U(T, M(T), j))$  is a maximum, where we take expectation, using our failure model, with respect to  $M(T)$  and  $j$ . This maximization is quite complex, and must be done numerically via computer. The details are found in the paper, but the end result is best displayed as a graph of time against expected utility (figure 4); in this case one can see that the time one should test the software for is about 3.5 units.

Expected utility

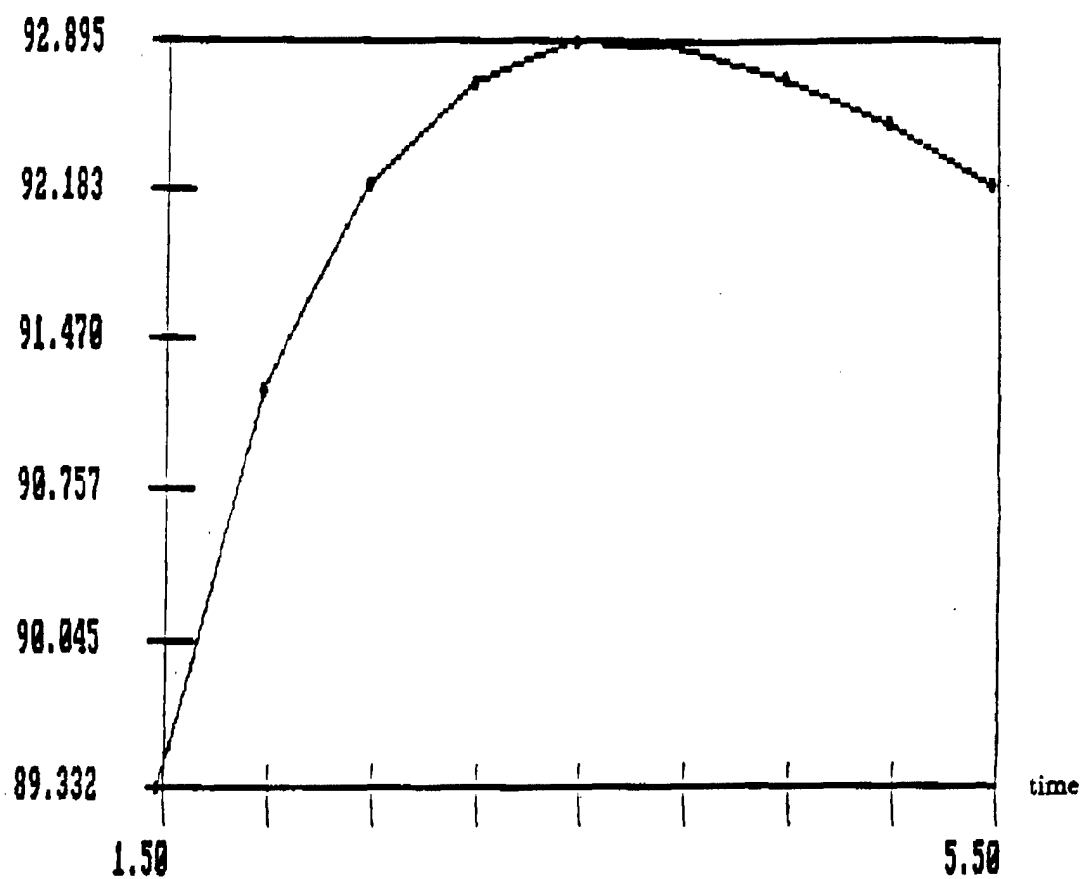


Figure 4 Time of testing versus expected utility for the model in Section 5

## 6. Conclusion.

This paper has attempted to review the main methods, and some of the more well known models, that have been used by the statistics community in the area of software reliability. The first models were almost always based on looking at the failure rate of the software: later on the idea of modeling number of failures by a Poisson process was used and then most recently auto-regressive processes have been suggested as an alternative to the failure rate method. Application of the failure models, such as to the optimal testing decision problem, is another important aspect to the field.

Earlier it was pointed out that there is almost no data on the reliability of commercial software, due to the sensitive nature of that information. A possible method of overcoming this problem would be to have more interaction between the statistics and computer science communities. In the future, such interaction seems essential if models are to become more realistic and useful, and it is perhaps surprising that there are so few links between the two groups today.

There still remains much to be researched in this field. In the case of optimal testing, plans for two-stage and sequential testing need to be developed, whilst the verification of current and future models is likely to remain a problem. Nevertheless, because of the increasing presence of computers in all aspects of our daily lives, the topic of software reliability can only become more important in the future.

### **Acknowledgements**

This report is based on a series of lectures given by the first author during the Fall of 1991 when he was the C. C. Garvin Endowed Visiting Professor of Computer Science and of Statistics at the Virginia Polytechnic Institute and State University. The comments of Professors I. J. Good, R. Krutchkoff, P. Palletas and K. Hinkelmann, and of students in the class, are gratefully acknowledged. We also thank Dr. Julia Abrahams of the Office of Naval Research for stimulating us to write this paper and for her comments on improving its readability. Also acknowledged are the inputs of Ms. Sylvia Campodonico and the assistance of Mr. Jingxian Chen.

## References

- Goel, A.L. (1983) A Guidebook for Software Reliability Assessment. Technical Report RADC-TR-83-176
- Goel, A.L. & Okomuto, K. (1978) An Analysis of Recurrent Software Failures on a Real-time Control System. Proc. ACM Annu. Tech. Conf., ACM, Washington D.C., pp. 496-500
- Goel, A.L. & Okomuto, K. (1979) Time-dependent Error Detection Rate Model for Software Reliability and other Performance Measures. IEEE Transactions on Reliability, vol. R-28, pp. 206-211
- Jelinski, Z. & Moranda, P. (1972) Software Reliability Research. Statistical Computer Performance Evaluation, W. Freiberger editor. New York: Academic, pp. 465-484
- Landberg, N. and Singpurwalla, N.D. (1985) A Unification of Some Software Reliability Models. SIAM J. Sci. Stat. Comput., vol. 6, no. 3, pp.781-790
- Littlewood, B. (1980) A Bayesian Differential Debugging Model for Software Reliability. Proceedings of IEEE COMPSAC
- Littlewood, B. & Verall, J.L. (1973) A Bayesian Reliability Growth Model for Computer Software. Applied Statistician, vol. 22, pp. 332-346
- Mazzuchi & Soyer (1988) A Bayes Empirical-Bayes Model for Software Reliability. IEEE Transactions on Reliability, vol. 37, no. 2, pp. 248-254
- Moranda, P.B. (1975) Prediction of Software Reliability and its Applications. Proceedings of the Annual Reliability and Maintainability Symposium, Washington D.C., pp. 327-332
- Musa, J.D. & Okomuto, K. (1984) A Logarithmic Poisson Execution Time Model for Software Reliability Measurement. Proceedings of the 7th International Conference on Software Engineering, Orlando, Florida, pp. 230-237

- Schick, G.J. & Wolverton, R.W. (1978) Assessment of Software Reliability. Proc. Oper. Res., Physica-Verlag, Wirzburg-Wien, pp. 395-422
- Singpurwalla, N.D. (1991) Determining an Optimal Time for Testing and Debugging Software. IEEE Transactions on Software Engineering, vol. 17, no. 4, pp. 313-319
- Singpurwalla, N.D. and Soyer, R. (1985) Assessing (Software) Reliability Growth Using a Random Coefficient Autoregressive Process and its Ramifications. IEEE Transactions on Software Engineering, vol. SE-11, no. 12, pp. 1456-1464



STATISTICAL METHODS APPLIED TO VOCATIONAL COUNSELING DATA  
OBTAINED FROM ARMY VETERANS

GENE DUTOIT  
DISMOUNTED WARFIGHTING BATTLE LABORATORY  
FORT BENNING, GEORGIA  
AND  
JOHN MOBLEY  
SKINNER & ASSOCIATES  
COLUMBUS, GEORGIA

ABSTRACT. THE SECOND AUTHOR OF THIS POSTER PRESENTATION IS UNDER CONTRACT WITH THE DEPARTMENT OF VETERAN AFFAIRS TO PROVIDE VOCATIONAL AND EDUCATIONAL COUNSELING TO ARMY MILITARY PERSONNEL WHO WILL BE LEAVING MILITARY SERVICE IN A SHORT TIME. THIS PARTICULAR COUNSELING SERVICE IS PROVIDED PRIMARILY TO THE INFANTRY BRANCH STATIONED AT FORT BENNING, GEORGIA. EACH SOLDIER IS GIVEN A BATTERY OF TESTS; APPTITUDE, ABILITY, INTERESTS, CAREER DEVELOPMENT AND PERSONALITY. THESE INSTRUMENTS COLLECTIVELY PROVIDE INFORMATION TO ASSIST THE COUNSELOR IN PROVIDING INDIVIDUAL GUIDANCE. THIS POSTER SESSION DID NOT FOCUS ON THE SCIENCE AND DECISION MAKING PROCESS OF INDIVIDUAL COUNSELING BUT EXAMINED AND SHOWED SOME OF THE RELATIONSHIPS THAT EXIST BETWEEN THE DIFFERENT PSYCHOMETRIC INSTRUMENTS. THESE STATISTICAL RELATIONSHIPS CAN BE USED TO GAIN INSIGHTS AND TO MAKE DECISIONS ABOUT THE VETERANS AS A GROUP AND TO PROVIDE THE VOCATIONAL COUNSELING COMMUNITY INFORMATION ABOUT THE VALIDITY OF THE INSTRUMENTS. BASIC STATISTICAL METHODS WERE USED TO ANALYZE THESE DATA. THE RESULTS OF FACTOR ANALYSIS WERE ESPECIALLY USEFUL FOR PROVIDING INSIGHTS ABOUT THE STRUCTURAL RELATIONSHIPS THAT EXIST BETWEEN THE DIFFERENT SCALES AND INSTRUMENTS. IN MANY CASES THE GRAPHICAL DISPLAYS SHOWED THAT THE TEST BATTERY CONSISTS OF A SET OF MUTUALLY SUPPORTING INSTRUMENTS.

THIS PAPER WILL ACCOMPANY THE POSTER PRESENTATION WHICH IS INCLOSED. THE READER SHOULD CONSIDER THE ABSTRACT (GIVEN ABOVE) AND EACH OF THE NINE POSTER DISPLAYS IN ORDER TO UNDERSTAND THE CONTENTS OF THE POSTER PRESENTATION. IF THERE ARE ANY QUESTIONS REGARDING INTERPRETATION OF THE POSTER DISPLAYS, PLEASE CALL THE FIRST AUTHOR AT (706)545-3165/3166 OR DSN 545-3165/3166. EACH POSTER DISPLAY WILL BE DISCUSSED IN ORDER.

POSTER DISPLAY 1. THIS DISPLAY SHOULD BE SELF-EXPLANATORY. NOTE THE PRIORITIES OF THE THREE GOALS. THE INDIVIDUAL SOLDIER COMES FIRST.

POSTER DISPLAY 2. THE SUBJECTS ARE DESCRIBED AS A GROUP.

POSTER DISPLAY 3. THIS IS A LISTING OF THE PSYCHOMETRIC TESTS (INSTRUMENTS) THAT WERE ADMINISTERED TO EACH SUBJECT / SOLDIER AS PART OF THE COUNSELING PROCEDURE. THIS DISPLAY TAKES TWO VIEWGRAPHS.

POSTER DISPLAY 4. THIS IS A SUMMARY OF THE RESPONSES OBTAINED FROM THE INVENTORY CALLED "MY VOCATIONAL SITUATION" (THE SECOND INSTRUMENT LISTED ON POSTER DISPLAY 3). THE ITEMS MARKED WITH AN "\*" ARE THOSE

RESPONSES THAT RESPONDENTS MARKED EITHER TRUE OR YES AT LEAST 50% OF THE TIME ( LEVEL OF SIGNIFICANCE OF 5% ). THIS WAS A SUBJECTIVE CRITERION FOR FOCUSING SOME CONCERN FOR THE COUNSELING PROCESS FOR THE SUBJECTS AS A GROUP. THE NEED TO FOLLOW UP THESE SIGNIFICANT RESPONSES WITHIN THE ARMY SYSTEM IS EXPRESSED AT THE BOTTOM OF THIS POSTER DISPLAY.

POSTER DISPLAY 5. THIS DISPLAY GIVES EXAMPLES OF THE CORRELATIONS AND SCATTER PLOTS BETWEEN SOME OF THE INSTRUMENTS AND SUB-SCALES LISTED ON POSTER DISPLAY 3. THE TOP FIGURE SHOWS THE PLOT BETWEEN INTELLIGENCE AND ABSTRACT REASONING ( $R=.9$ ) AND THE BOTTOM FIGURE IS THE PLOT BETWEEN EXTRAVERSION AND INTROVERSION ( $R=-.92$ ). THESE PARTICULAR RELATIONSHIPS WERE EXPECTED AND DESIRED FOR VALID MEASUREMENTS OF THESE PSYCHOLOGICAL DIMENSIONS. THE CORRELATIONS BETWEEN ALL THE SCALES ARE EXPLORED FURTHER IN POSTER DISPLAYS 6,7 AND 8.

POSTER DISPLAY 6. THIS IS THE COMPLETE CORRELATION MATRIX FOR ALL TWENTY SCALES EVALUATED FOR EACH SUBJECT. THIS MATRIX WAS PREPARED FOR INPUT TO A FACTOR ANALYSIS ROUTINE.

POSTER DISPLAY 7. THIS GIVES SOME INTERPRETATIONS OF THE FACTOR ANALYSIS. THIS DISPLAY IS INTENDED TO BE SELF-EXPLANATORY. THE SIX FACTORS THAT WERE EXTRACTED ARE DESCRIBED/INTERPRETED ON THE NEXT DISPLAY.

POSTER DISPLAY 8. THE FACTOR LOADINGS FOR ALL TWENTY SCALES ARE GIVEN HERE. THESE ARE CLASSICAL TEXT-BOOK RESULTS. THE LOADINGS ARE "AS EXPECTED" FOR ALL SIX FACTORS. IN A SENSE THIS CAN BE INTERPRETED AS HELPING TO CONFIRM THE VALIDITY OF EACH INSTRUMENT AND SUB-SCALE FOR USE IN COUNSELING THE TARGET GROUP OF SUBJECTS.

POSTER DISPLAY 9. THIS SUMMARY DISPLAY WRAPS UP THE MAJOR FINDINGS SHOWN AND DISCUSSED IN THE PREVIOUS VIEWGRAPHS.



## THE SITUATION

- PROVIDE VOCATIONAL & EDUCATIONAL COUNSELING TO DEPARTING ARMY PERSONNEL
- FOCUS ON INFANTRY BRANCH AT FT BENNING, GA
- EACH SOLDIER IS GIVEN A BATTERY OF TESTS

### GOALS...USEFUL FEEDBACK TO:

THE ARMY SOLDIER AS AN INDIVIDUAL  
THE ARMY IN GENERAL  
THE COUNSELING COMMUNITY

## POSTER DISPLAY 2

### SUBJECTS

- 87 MALES; ARMY MILITARY, COUNSELING PRIOR TO LEAVING SERVICE
- MOST ARE LEAVING BECAUSE OF THE MILITARY DRAW-DOWN
- A 50% SAMPLE OF THE AVAILABLE SUBJECTS AT SOME POINT IN TIME
- NCOs & OFFICERS HAVE BEEN POOLED (DATA COLLECTED THAT WAY)
- BRANCH IS GENERALLY INFANTRY • FT BENNING GA
- AVG AGE = 33.8 YRS
- STD DEV = 8.2 YRS
- THE DISTRIBUTION PLOT OF AGE IS SKEWED TOWARD THE OLDER SUBJECTS
- RANGE OF AGE = 20 YRS TO 52 YRS
- LILLIEFORS TEST INDICATES NORMALITY OF AGE ( $P = .0593$ )..... "WEAK DECISION"
- NO 'AGE' OUTLIERS DETECTED IN THE BOXPLOTS (POS SKEW WAS VISIBLE)

NOTE 1 THIS WAS AN 'AFTER THE FACT' DATA ANALYSIS. SOME HOLES DO EXIST

NOTE 2 THE SKEW IN THE DIRECTION OF AGE WAS NOT UNEXPECTED.

THE OLDER SUBJECTS ARE, GENERALLY, MORE ELIGIBLE TO RETIRE  
AND THIS DISTRIBUTION REFLECTS THE ARMY'S POLICY TO RELEASE  
OR RETIRE THE OLDER SOLDIERS.

## INSTRUMENTS

1. "THE SELF DIRECTED SEARCH"(INTERESTS)
  - A. REALISTIC
  - B. INVESTIGATIVE
  - C. ARTISTIC
  - D. SOCIAL
  - E. ENTERPRISING
  - F. CONVENTIONAL
2. "MY VOCATIONAL SITUATION"(CAREER DEVELOPMENT)

BASED ON THE ASSUMPTION THAT MOST DIFFICULTIES IN VOCATIONAL DECISION MAKING FALL INTO ONE OR MORE OF THE FOLLOWING CATEGORIES: PROBLEMS OF VOCATIONAL IDENTITY; LACK OF INFORMATION ABOUT JOBS OR TRAINING; OR ENVIRONMENTAL OR PERSONAL BARRIERS.
3. "SHIPLEY INSTITUTE OF LIVING SCALE"(APTITUDE)
  - A. VERBAL
  - B. ABSTRACT
  - C. INTELLEGE
4. "CAREER PLANNING PROGRAM"(ABILITY)
  - A. READING
  - B. SPATIAL RELATIONS
  - C. NUMERICAL SKILLS
  - D. MECHANICAL REASONING

## **INSTRUMENTS (CONT)**

### **5. "INTRODUCTION TO TYPE" - BRIGGS/MYERS (PERSONALITY)**

- A. EXTROVERSION**
- B. INTROVERSION**
- C. SENSING**
- D. INTUITION**
- E. THINKING**
- F. FEELING**
- G. JUDGING**
- H. PERCEIVING**

# POSTER DISPLAY 4

## INFORMATION ABOUT THE VOCATIONAL NEEDS OF THE SOLDIER

### my vocational situation

Name: \_\_\_\_\_ Date: \_\_\_\_\_  
Education completed: \_\_\_\_\_ Other: \_\_\_\_\_

List all the occupations you are considering right now:

Try to answer each of the following statements as mostly TRUE or mostly FALSE. Circle the answer that best represents your present opinion.

In thinking about your present job or in planning for an occupation or career: (see p. 10 for TRUE)			$\hat{p}$
	T	F	
1. I need reassurance that I have made the right choice of occupation.	T	F	.61
2. I am concerned that my present interests may change over the years.	T	F	.55
3. I am uncertain about the occupations I could perform well.	T	F	.52
4. I don't know what my major strengths and weaknesses are.	T	F	.48
* 5. The jobs I can do may not pay enough to live the kind of life I want.	T	F	.63
6. If I had to make an occupational choice right now, I am afraid I would make a bad choice.	T	F	.34
* 7. I need to find out what kind of career I should follow.	T	F	.75
* 8. Making up my mind about a career has been a long and difficult problem for me.	T	F	.59
9. I am confused about the whole problem of deciding on a career.	T	F	.39
10. I am not sure that my present occupational choice or job is right for me.	T	F	.45
* 11. I don't know enough about what workers do in various occupations.	T	F	.67
12. No single occupation appeals strongly to me.	T	F	.46
13. I am uncertain about which occupation I would enjoy.	T	F	.49
* 14. I would like to increase the number of occupations I could consider.	T	F	.82
15. My estimates of my abilities and talents vary a lot from year to year.	T	F	.34
16. I am not sure of myself in many areas of life.	T	F	.31
17. I have known what occupation I want to follow for less than one year.	T	F	.28
18. I can't understand how some people can be so set about what they want to do.	T	F	.29

(over) ☐

For questions 19 and 20, circle YES or NO

19. I need the following information:		$\hat{p}$
* How to find a job in my chosen career	Y N	.06
What kinds of people enter different occupations	Y N	.55
* More information about employment opportunities	Y N	.89
* How to get the necessary training in my chosen career	Y N	.74

Other: \_\_\_\_\_

20. I have the following difficulties:

I am uncertain about my ability to finish the necessary education or training.	Y	N	.38
I don't have the money to follow the career I want most.	Y	N	.53
I lack the special talents to follow my first choice.	Y	N	.34
An influential person in my life does not approve of my vocational choice.	Y	N	.12

Anything else? \_\_\_\_\_

Other comments or questions: \_\_\_\_\_

Developed by John L. Holland, Dorcas C. Dwyer, and Paul G. Power.  
© 1980 by John L. Holland, Dorcas C. Dwyer, and Paul G. Power. All Rights Reserved. No reproduction of this material is authorized without written permission of the Publisher.

Published by Consulting Psychologists Press, Inc., Palo Alto, CA

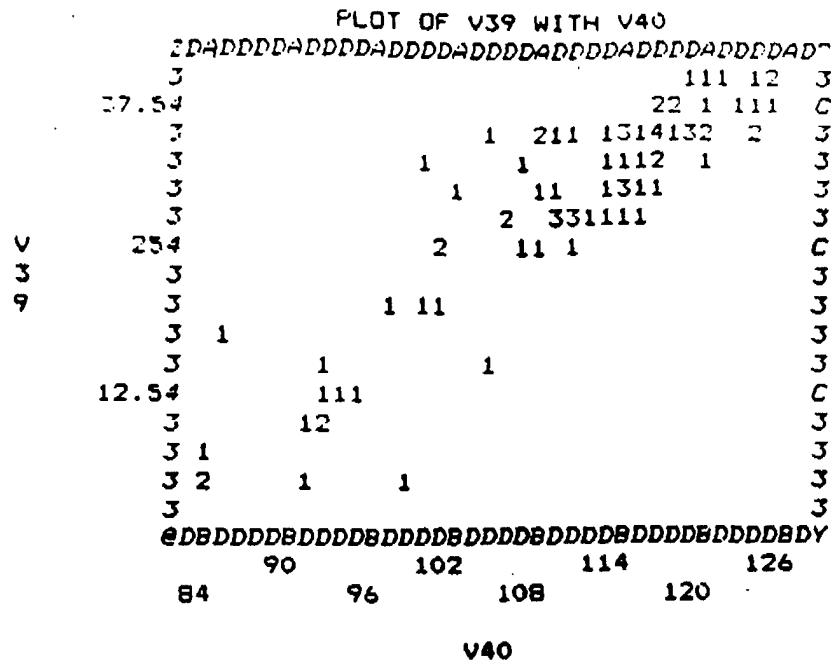
## A NEED TO FOLLOW UP THE "SIGNIFICANT RESPONSES" OF CONCERN:

- HELPING SOLDIERS MAKE CAREER DECISIONS
- EXPLAINING THE CAREER OPTIONS
- INFORMATION ABOUT OPPORTUNITIES AND TRAINING

# POSTER DISPLAY 5

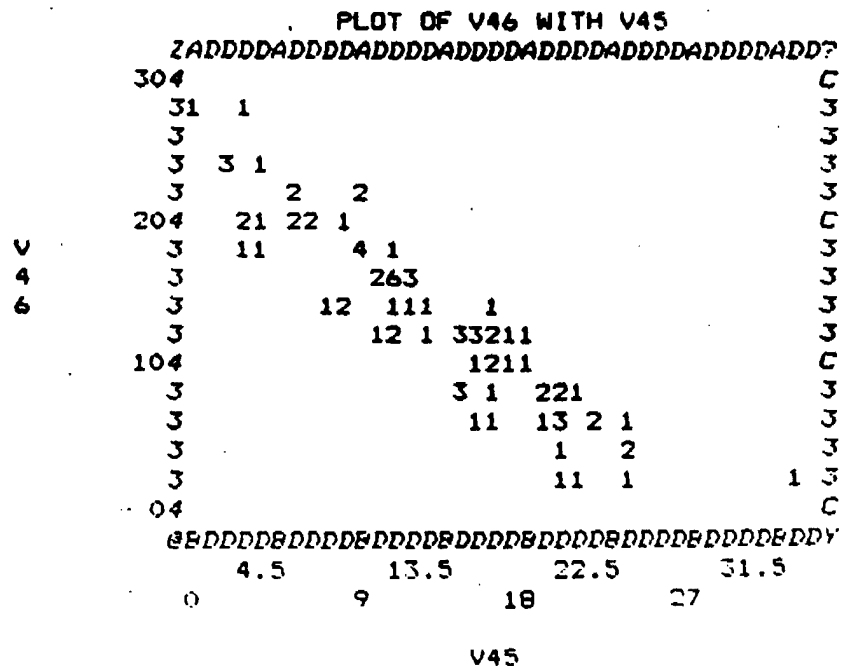
Correlations: V40

V37	<u>intelligent</u>	-.3712**
V38	verbal	.7402**
V39	abstract	.8959**
V40		1.0000**
V41	reading	-.6232**
V42	spatial	.5656**
V43	numerical	.6742**
V44	mechanical	.6825**
V45		-.0128
V46		.0488
V47		-.2561*
V48		.1978
V49		.1140
V50		-.0665
V51		-.1128
V52		.0767



Correlations: V45

V36	<u>extroversion</u>	-.1477
V37		-.0392
V38		-.0560
V39		-.0261
V40		-.0128
V41		.0752
V42		.0032
V43		.0305
V44		.0859
V45		1.0000**
V46	<u>introversion</u>	-.9232**
V47		-.2569*
V48		.2282
V49		-.0574
V50		.0693
V51		.2671*
V52		-.2110





# DISPLAY POSTER 6

## CORRELATION MATRIX FOR ALL SCALES

### INPUT TO THE FACTOR ANALYSIS

#### - - - - FACTOR ANALYSIS - - - -

Analysis Number 1 Matrix input

Correlation Matrix:	INTELLECTUAL X1	ARTIST X2	SOCIAL X3	ENVIRONMENTAL X4	CONVENTIONAL X5	VERBAL X6	ABSTRACT X7
X1	1.00000						
X2	.47000	1.00000					
X3	.38000	.46000	1.00000				
X4	.43000	.34000	.62000	1.00000			
X5	.38000	.21000	.42000	.67000	1.00000		
X6	.26000	.11000	-.04000	.06000	.09000	1.00000	
X7	.24000	.08000	.09000	.11000	.08000	.56000	1.00000
X8	.31000	.14000	.08000	.13000	.08000	.74000	.90000
X9	.13000	.09000	-.03000	.21000	.23000	.62000	.58000
X10	.29000	.08000	.01000	-.01000	-.03000	.44000	.54000
X11	.19000	.04000	-.05000	-.01000	.05000	.53000	.61000
X12	.29000	.09000	-.02000	.10000	-.03000	.64000	.63000
X13	.17000	.22000	.40000	.44000	.22000	-.06000	-.03000
X14	-.24000	-.19000	-.44000	-.49000	-.31000	.11000	.07000
X15	-.32000	-.36000	-.18000	-.15000	.16000	-.21000	-.18000

Page 4

SPSS/PC+

7/17/92

#### - - - - FACTOR ANALYSIS - - - -

	X1	X2	X3	X4	X5	X6	X7
X16	.29000	.39000	.16000	.16000	-.15000	.25000	.05000
X17	.05000	.08000	-.09000	.04000	.09000	.12000	.13000
X18	.12000	-.07000	.10000	.02000	.04000	-.06000	-.09000
X19	.03000	.04000	.34000	.14000	.25000	-.17000	-.04000
X20	.01000	-.03000	-.25000	-.03000	-.16000	.04000	.01000
	INTELLECTUAL X8	READING X9	SPIRITUAL X10	NUMERICAL X11	MECHANICAL X12	ENVIRONMENTAL X13	IMAGINATION X14
X8	1.00000						
X9	.62000	1.00000					
X10	.57000	.55000	1.00000				
X11	.67000	.64000	.68000	1.00000			
X12	.68000	.62000	.52000	.56000	1.00000		
X13	-.01000	.08000	.01000	.03000	.09000	1.00000	
X14	.05000	-.04000	.01000	-.01000	.04000	-.92000	1.00000
X15	-.26000	-.05000	-.25000	-.17000	-.16000	-.26000	.22000
X16	.20000	.07000	.20000	.13000	.15000	.23000	-.17000
X17	.11000	.17000	.14000	.16000	.09000	-.06000	.11000

Page 5

SPSS/PC+

7/17/92

#### - - - - FACTOR ANALYSIS - - - -

	X8	X9	X10	X11	X12	X13	X14
X18	-.07000	-.07000	-.08000	-.15000	-.03000	-.07000	-.11000
X19	-.11000	.01000	.02000	.02000	-.07000	.26000	-.27000
X20	.08000	-.01000	-.01000	-.04000	.01000	-.21000	.19000
	SENSING X15	INTUITION X16	THINKING X17	FEELING X18	IMAGINE X19	PERCEIVING X20	
X15	1.00000						
X16	-.85000	1.00000					
X17	.05000	-.05000	1.00000				
X18	-.03000	.10000	-.70000	1.00000			
X19	.26000	-.27000	.12000	-.11000	1.00000		
X20	-.24000	.22000	-.10000	.12000	-.90000	1.00000	

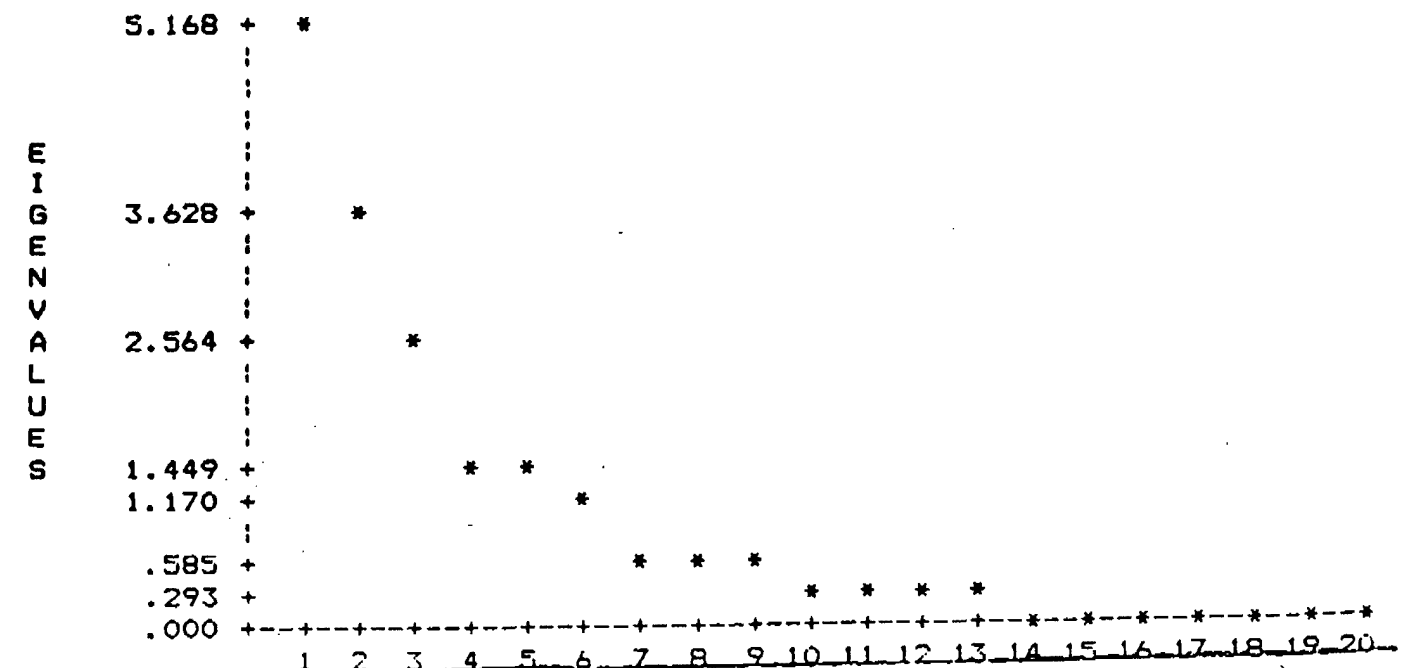
# POSTER DISPLAY 7

## INTERPRETATION OF THE FACTOR ANALYSIS

1. CORRELATION MATRIX INPUT  
20 VARIABLES; X1 TO X20  
N=87 CASES  
43 PAIRWISE CORRELATIONS GREATER THAN .30
2. BARTLETT TEST OF SPHERICITY INDICATES THAT THE CORRELATION MATRIX IS NOT AN IDENTITY MATRIX. FAVORABLE FOR FACTOR ANALYSIS.
3. THE KAISER-MEYER-OLKIN MEASURE OF SAMPLING ADEQUACY (.63) IS A MEDIOCRE VALUE. ANOTHER MEASURE OF SAMPLING ADEQUACY IS SOMEWHAT FAVORABLE TO FACTOR ANALYSIS.
4. BASED ON AN EIGENVALUE CRITERION OF "AT LEAST ONE" AND THE SCREE PLOT, THERE ARE SIX FACTORS INVOLVED WITH THESE VARIABLES. A VARIMAX ROTATION WAS SELECTED. THE FINAL FACTOR STATISTICS ARE PRESENTED BELOW:

Factor	Eigenvalue	Pct of Var	Cum Pct
1	5.16778	25.8	25.8
2	3.62840	18.1	44.0
3	2.56355	12.8	56.8
4	1.62856	8.1	64.9
5	1.44906	7.2	72.2
6	1.16989	5.8	78.0

## - - - - FACTOR ANALYSIS - - - -



## POSTER DISPLAY 8

### 5. FACTOR 1 ( ALL + LOADINGS)

X8....INTELLEGEENCE, SHIPLEY  
X7....ABSTRACT, SHIPLEY  
X11...NUMERICAL SKILLS, CAREER PLANNING PROGRAM  
X12...MECHANICAL REASONING, CAREER PLANNING PROGRAM  
X9....READING, CAREER PLANNING PROGRAM  
X6....VERBAL, SHIPLEY  
X10...SPATIAL, CAREER PLANNING PROGRAM

COMMENT: THIS FACTOR ACCOUNTS FOR BOTH THE SHIPLEY AND THE  
CAREER PLANNING MEASURES. THIS IS PROBABLY INTELLEGEENCE.  
INDICATES CONSTRUCT VALIDITY.

### FACTOR 2 ( ALL + LOADINGS)

X5....CONVENTIONAL, SELF DIRECTED SEARCH  
X4....ENTERPRISING, SELF DIRECTED SEARCH  
X3....SOCIAL, SELF DIRECTED SEARCH  
X1....INVESTIGATIVE, SELF DIRECTED SEARCH

COMMENT: ALL LOAD ON THE SAME INSTRUMENT WHICH CLAIMS TO  
DETERMINE INTERESTS. INDICATES CONSTRUCT VALIDITY.

### FACTOR 3

X15...SENSING(ORDERLY), (-), INTRODUCTION TO TYPE  
X16...INTUITION, (+), INTRODUCTION TO TYPE  
X2....ARTISTIC, (+), SELF DIRECTED SEARCH

COMMENT: LOADED POSITIVELY ON ARTISTIC AND INTUITION AND NEGATIVELY  
ON THE ORDERLY TYPE. THIS MEETS A PERCEIVED IMAGE OF THE  
ARTISTIC TYPE OF PERSONALITY. CONSTRUCT VALIDITY.

### FACTOR 4

X13...EXTRAVERSION, (+), INTRODUCTION TO TYPE  
X14...INTROVERSION, (-), INTRODUCTION TO TYPE

COMMENT: THESE TWO LOADINGS ARE STRONGLY CORRELATED (NEGATIVE)  
CALL IT A FACTOR OF EXTRAVERSION. CONSTRUCT VALIDITY.

### FACTOR 5

X20...PERCEIVING, (-), INTRODUCTION TO TYPE  
X19...JUDGING, (+), INTRODUCTION TO TYPE

COMMENT: SAME COMMENTS AS FACTOR 4 ABOVE. CALL IT A FACTOR OF  
JUDGING. CONSTRUCT VALIDITY.

### FACTOR 6

X18...FEELING, (-), INTRODUCTION TO TYPE  
X17...THINKING, (+), INTRODUCTION TO TYPE

COMMENT: SAME COMMENTS AS FACTOR 4 ABOVE. CALL IT A FACTOR OF  
THINKING. CONSTRUCT VALIDITY.



## SUMMARY

### CONCERNING THE SOLDIER

- NEED TO HELP SOLDIERS IN MAKING CAREER DECISIONS  
MORE KNOWLEDGE TO GET MORE SELF CONFIDENCE
- MORE INFORMATION ABOUT OPPORTUNITIES & TRAINING

### THE COUNSELING COMMUNITY

- TRENDS WITH RESPECT TO AGE ARE IN THE EXPECTED DIRECTION
- 'INTELLEGENGE' CORRELATES WITH EXPECTED VARIABLES
- MANY VARIABLES ARE CORRELATED IN THE EXPECTED DIRECTION
- EXPECTED CONFIRMATORY FACTOR ANALYSIS RESULTS OBTAINED

# The MDL Principle - A Tutorial

J. Rissanen

IBM Almaden Research Center, San Jose, Ca 95120-6099

**Abstract:** The MDL (Minimum Description Length) principle is meant to provide guidance to inductive inference and modeling by posing them literally as data compression problems. It is seen to be a direct generalization of the Least Squares and the Maximum Likelihood techniques as well as the Maximum Entropy principle. It also provides a concrete code length based interpretation of the priors in Bayesian inference, and it permits their optimization in the light of the data. Finally, the shortest code length generalizes Shannon's information by inclusion of a term that accounts for the effect of estimation. A basic result in the MDL theory generalizes Shannon's fundamental noiseless coding theorem and sets bounds to the main data processing tasks of data compression, estimation, and prediction.

## 1. INFERENCE PRINCIPLES

Inductive inference is the familiar process aimed at extrapolating general laws from a given set of data generated by some physical machinery or, as put by Maxwell, it is the process for finding the 'go' of it. This, of course, is also the way to learn from experience, for since in general the current data will not occur exactly in the future it is the summary information represented by the laws that we can learn. After all, the storage and recalling capacities of the brain would rapidly be overwhelmed if we tried to store all the data we perceive. Despite the common belief in a mystical 'true' law, which obviously is a mathematical and hence linguistic concept, there is no unique way to construct such an extrapolation. What is even worse, it is impossible to formalize the induction problem with a perfect inference scheme as the solution. Indeed, the definitive manifestation of a found

law is that it would predict optimally the future observations. But this would make the law dependent on future data, which we do not have today, and because any law we find must be determined by the given current data, we arrive at a contradiction. Squeezed between such conflicting demands we must settle for less and ask merely for a principle to select an extrapolation law, or perhaps less ambitiously a *model* of the data generating machinery, which has intuitive appeal and, more importantly, which provides good models and model classes for new and nontrivial problems. The key requirement here is that the principle, or the model selection criterion, should have a meaningful data dependent interpretation. A further bonus would be if the principle could be proved to have various desirable mathematically defined properties in the cases where analysis is possible.

In this paper we study the MDL (Minimum Description Length) principle for model selection. Expressed in broad terms, the principle calls for that model class or model, as the case may be, with which the observed data can be encoded with the fewest number of binary digits. In this, it is important that the optimal model itself needed to do the job is also included. Despite such a 'nonstatistical' enunciation of the principle, it actually is a direct extension of the line of the most important inference principles of them all, beginning with the idea of the least squares by Johan Lambert over 200 years ago. It was recast by Gauss as maximization of the distribution bearing his name, and it was developed into the general but still 'local' Maximum Likelihood principle by Fisher. In fact, the code length defines a probability somewhat analogously to the way the squared deviations define a normal distribution, and the MDL principle can equally well be called a 'global' Maximum Likelihood principle to emphasize the fact that any two models or model classes may be compared, regardless of their type and the number of parameters in them.

The MDL principle is also related to but is distinct from Bayesian inference, which at least in its original form is based upon Bayes' theorem. This theorem transforms an initial distribution on the parameters, assumed to express prior knowledge about the 'true' parameter value, in the light of the observed data into the more informative posterior distribution. This, in turn, can either be used for estimation

of the 'true' parameter value or to optimize a suitable risk function of the future performance and to make intelligent decisions. The weak spot in this in itself sensible reasoning is the controversial initial 'prior' distribution, its meaning and choice, as well as the extent to which the nebulous 'prior knowledge' can be expressed as a distribution on parameters which themselves are artifacts in more or less arbitrarily chosen models. In the MDL formalism anything that can be described in a finite number of distinct symbols, which certainly includes parameters and data, will get a code length induced probability, which then will be the concrete meaning of probabilities on parameters. But more importantly, the MDL principle imposes a restriction on the prior probabilities, which permits their optimization without paradoxes and hence makes the dream of the Bayesians, the so-called empirical data fitted priors, come true. The restriction comes from the simple requirement that since an object, say the integer  $n = 3729$ , cannot be described in a prefix manner (see Appendix) with fewer than about  $\log n + 2 \log \log n$ , or 20 bits, using commonly agreed ways of encoding, it is impossible in the MDL framework to assign to this number a prior probability larger than about  $2^{-20}$ . To put it differently, a Bayesian might have on good authority the piece of prior information that the probability of the given integer is  $1/2$ , and there is nothing in the Bayesian inference contradicting this belief. This would allow encoding of the integer with  $\log 2 = 1$  bit. However, in the MDL framework this information must be described to others, which means that about 20 bits are required to describe the special integer which has such a high probability as  $1/2$ , and nothing is gained. The point here is that in the MDL formalism description of objects should be done by universally available means using a natural language and conventional mathematics. If special means are desired, they must be explained; ie, redescribed, in the generally available terms. This has first the implication that different objects require different amounts of bits to describe them, which is determined by the way languages, including mathematics, are formed. In fact, that is how we distinguish between simple and complex things. Further, the nature of prior knowledge in this framework is something that is shared and generally known, which is in contrast with 'private' prior knowledge that some Bayesians subscribe to. We may, of course, still use special prior knowledge about the type of models we wish to fit to a particular set of data, knowledge generated

by others who may have studied a similar problem and which is not in the data we have. But such knowledge should not compete with the evidence; ie, the data, and if it does the data win! After all, in this game prior knowledge is supposed to help in explaining the existing data, rather than imagined nonexistent 'future' data.

Another but still partial view of the MDL principle is to regard it as some sort of a practical implementation of the ideas in the theory of algorithmic or Kolmogorov complexity, Solomonoff (1964), Kolmogorov (1965), and others. Indeed, if we regard a program with which a universal computer can generate the observed data, represented as a binary string, as its model, then the shortest program may be taken as the best model of the data. Although nonunique such a program must represent all the regular features in the string that on the whole can be expressed in the programming language for the machine, which clearly is the paramount requirement of a model. The unique length of the shortest programs for the string is called the Kolmogorov complexity of the string, and with the restriction that no program, regarded as a binary string, is a prefix of another, Chaitin (1975), the complexity defines a sort of universal prior distribution for the integers. Since parameters, truncated to a finite precision, may easily be encoded as integers, the central problem nagging the Bayesians seems to get solved, see Li and Vitanyi (1992). The fly in the ointment, however, is that the Kolmogorov complexity is not computable, except by approximations from above without our being able to form an adequate idea of the error. What is worse, the universal prior cannot even be approximated from either side, except if left unnormalized (the sum differing from unity), and despite the asymptotic universality properties of such an unnormalized 'prior', the hardly surprising conclusion is that it does not provide any help in tackling the inductive inference problems arising in practice.

How well does the MDL principle satisfy the above stated goals of intuitive appeal, utility, and analytic tests? Perhaps because of the intuitive idea of a model to be a short summary of the data, and the general feeling that redundancy is something to be avoided, many people find the code length minimization appealing - even those who are not familiar with coding theory. However, there are others



who are not convinced and frequently raise questions like 'Why is the code length criterion any good for selecting models, unless the application of the model is for data compression?'. Moreover, since the objective is to get a model that performs well in future data, rather than being able to compress the current data, one wonders why we shouldn't minimize an expected value of some such desired quantity as the prediction error. We dispose first of the second type of criterion, which actually forms the very foundation on which traditional statistics is erected. The required expectation presupposes the existence of a 'true' and unique underlying distribution so that the expected quantity, say the quadratic error, can be approximated from its samples by the appropriate estimation procedure for such a distribution. In other words, the real criterion that gets minimized will be a function of the current data, determined by the particular estimation procedure. The trouble is that this procedure and hence the result depend on the assumed 'true' distribution, which is anything but unique. For example, if we fit a polynomial curve to a set of data pairs and measure the error by the quadratic deviations, we are implicitly assuming a gaussian distribution with the mean defined by a polynomial of unknown degree. The higher the degree we pick the better fit we get; ie, the smaller the estimated mean square error, which is absurd.

A widely accepted guidance to avoid such absurdities is obtainable by analysis. Provided that there is a 'true' distribution generating the data, it is in some cases possible to deduce that the estimated mean performance, such as the mean prediction error, has both a bias and variance. Since the former gets smaller and the latter increases as the number of parameters increases we may seek a compromise by minimizing an appropriately weighted sum of the two effects. While adding the variance into the picture prevents models of extreme complexity from being optimal, the assumption of a 'true' distribution is untenable and deprives the criterion any data dependent interpretation. For the same reason, the choice of the degree of the compromise must be left for judgement, which means that we no longer have a rational basis to prefer one model over another, in particular when they are of different type having different numbers of parameters. The inevitable conclusion is that we cannot replace an intuitively appealing data dependent criterion by an esti-

mate of a dreamed 'ideal' criterion and hope to overcome the fundamental dilemma in inductive reasoning. In the MDL criterion the role of 'bias' and variance are played by the code lengths for the data and the model, the former getting smaller and the latter growing with an increasing number of parameters, so that the MDL principle strikes an automatic balance between the two terms without any arbitrary weighting factors.

We return to the first question, the intuitive appeal of the code length criterion to be minimized. We already mentioned its equivalent interpretation as a global maximum likelihood principle, which to us seems as the single most powerful way to assess goodness of models. It is easy to visualize a model with a highly peaked distribution, centered on the current data, as being able to provide a good predictor with few surprises, while in contrast a flat distribution constrains the data only weakly with plenty of room for deviations from any predicted behavior. In fact, we know of only two data dependent criteria, the prediction errors and the code length or, equivalently, probability, and they both will be optimized by the MDL model; see Rissanen (1984) and Weinberger et al (1992).

The chapter on the utility of the MDL criterion is, of course, not closed. What we can report is a steadily growing number of nontrivial successful applications, some of which are discussed in Gao and Li (1988), Hannan and Rissanen (1988), Leclerc (1989), Quinlan and Rivest (1989), Rissanen and Ristad (1992), Sheinvald et al (1992), Wax and Ziskind (1989), Dengler (1990), Rao et al (1991). In addition, the majority of the most successful order determination criteria for regression and time series problems discussed in the literature actually admit a code length interpretation. Finally as to the provable properties, all the versions of the MDL criteria have been shown to provide consistent estimates of the number of parameters as well as of their values in the usual analyzable model classes, Gerencser (1989), Hannan et al (1989), Hemerly and Davis (1989), Yu (1990). In the particular case of the linear least squares problems, the predictive MDL estimates extend the classical optimality properties of the unbiased least squares estimates to the estimation of the number of parameters as well as to the optimality of the mean accumulated

prediction errors, Rissanen (1986b), and others.

To conclude this introductory section we mention a conceptually important contribution of the marriage between the algorithmic and stochastic complexity theories. It is the fact that if the model classes we wish to contemplate include all the computable models, then there is no algorithm that will produce the MDL model as the output, when the observed data sequence is given as the input. This means that while the code length yardstick does permit comparison of any two models, we cannot form by computable means an assessment of how close to the optimal any model we have found is. This is the fundamental obstacle in all statistical work, which no amount of hypothesis testing nor anything else can overcome. On the positive side, however, this is the *only* fundamental obstacle; we can at least tackle the others, such as the estimation of the shortest code length, relative to a complex model class, which, to be sure, can be difficult enough. But, we feel, it is better to know one's 'enemy' rather than to be oblivious to it.

## 2. CODING WITH MODEL CLASSES

As one can argue, Rissanen (1989), virtually all models can be taken as probability distributions for the data of the two types,  $P(x^n|\theta)$  or  $P(y^n|x^n, \theta)$ , where  $x^n = x_1, \dots, x_n$  and  $y^n = y_1, \dots, y_n$  denote the data sets of any kinds of 'symbols' and  $\theta = \theta_1, \dots, \theta_k$  denotes the parameters of any kind and number  $k$ . Usually, the parameter values range over the real line while the data symbols range over a finite or a countable set. A few clarifying points might be in order to substantiate the made sweeping statement. First, by models in this context we mean the models that we actually fit to the data, rather than some mathematical abstractions such as those defined by ordinary or partial differential equations. These, of course, can be highly useful in suggesting good models of the kind we end up fitting to the data. Secondly, the distributional models may have any number of mathematical equalities and relations as parts; indeed, the probabilities involved often refer to the deviations from the deterministic 'laws' defining the model of interest, and, in

fact, they may get defined by the way the deviations are measured. The reason why we cannot separate such deviations at the outset as 'noise' is that they obviously get defined by whatever deterministic behavior we have selected to represent the model. Clearly, we may wish to apply our prior knowledge and model the 'noise', say the measurement errors, knowing the properties of the instrument, differently from the 'smooth' signal, but all this is included in the above formulation. Such a formulation of the notion of a model, which may appear to some as too vague, is just a testimony of the tremendous generality of the MDL principle and the ideas involved.

The central question which must be dealt with in order to make the MDL principle to work in applications is how to estimate the shortest code length with which the data can be encoded when a class of models is given. For this we need the basic results of Shannon's coding theory, see the appendix. They may be summarized by the single statement that the best way to encode data  $x^n$ , obtained from a single distribution  $P(x^n)$  by sampling, is to design a code such that this data set gets encoded with  $-\log P(x^n)$  binary digits. Clearly, since the code length must be integer-valued we can achieve the ideal to within one bit. We ignore the difference and call the quantity  $-\log P(x^n)$ , usually called the (self) *information*, the *ideal* code length for the string  $x^n$  under the stipulated conditions. Hence, we conclude already now that we can replace the knowledge of the 'true' distribution by the best code for the data - provided we know how to design it! Indeed, if under the agreed conditions we have been able to construct the shortest code, its length  $L(x^n)$  for the data also defines the largest probability  $2^{-L(x^n)}$  we have managed to assign to them, and hence it provides the best model of the above stated distributional kind.

Assume then that a set of models  $\mathcal{M}_k = \{P(x^n|\theta, \alpha)\}$  is selected for the data. The parameter  $\alpha$  is an integer-valued index, or a set of them, while  $\theta = \theta_1, \dots, \theta_k$  consists of components, which range over various subsets of the real numbers. To keep things simple we ignore the parameter  $\alpha$ , which anyway usually requires an order of magnitude fewer bits than the real-valued parameters  $\theta$ . These, to be sure, themselves are truncated in order to admit encoding with a finite code length. The

number of parameters  $k$  in  $\theta$  is often itself variable to be determined optimally. In such a case the model class of interest is  $\mathcal{M} = \bigcup_k \mathcal{M}_k$ . Notice that even the so-called nonparametric models are actually included; after all, when fitting such models we end up fitting a lot of parameters, sometimes even of the order of  $n$ .

There are three basic ways to encode data relative to the model class  $\mathcal{M}_k$ , each of which has its advantages and disadvantages. After all, encoding data in such circumstances does not admit such a clear-cut solution as given by Shannon's theory. We outline all three methods.

### 2.1. Two-Part Coding

We begin with the most general method, the so-called two-part coding, which is intuitive and often simple to apply. Each parameter value  $\theta$  specifies a distribution, which by Shannon's work permits encoding of the data with about  $L(x^n|\theta) = -\log P(x^n|\theta)$  bits. However, decoding can only be done if the decoder knows the parameter value which the encoder used. Hence, we need a preamble in the total code to specify the chosen parameter. And since we are not allowed to use a comma to separate the preamble from the rest, the code for the parameters must be a prefix code, see the appendix. It is clear that to encode the parameters by a finite binary string they must be truncated to a finite precision. If we take the precision the same  $\delta = 2^{-q}$  for all of them, we can represent  $\theta_i$  with the largest integer multiple of the precision, not exceeding  $\theta_i$ , written as  $\lfloor(\theta_i|2^q)$ . Hence, the number of bits needed for each parameter is about  $\log(|\theta_i|2^q)$ . Actually, because of the prefix requirement, a few more bits are needed, however, not more than  $2\log\log(|\theta_i|2^q)$  bits, (see Appendix), which we ignore as well as the sign bit for the parameter and the code length for the integer  $q$  itself. We then can encode the data with about

$$L(x^n|\mathcal{M}_k) = \min_{\theta, q} \{-\log P(x^n|2^{-q}\lfloor(|\theta|2^q)) + kq + \sum_{i=1}^k \log(|\theta_i|)\}, \quad (2.1)$$

where we wrote  $\lfloor(|\theta|2^q)$  as the vector of the components  $\lfloor(|\theta_i|2^q)$ . Notice that an increase of the value of  $q$  increases the second term  $kq$  but reduces the first in the

worst case, when the truncated parameter deviates maximally from the unrestricted minimizing value  $\hat{\theta}$ . Hence, there is an optimum worst case precision which can be found numerically. In particular, the second and the third terms may be defined to be the optimal model complexity

$$L(\hat{\theta}, \hat{q} | \mathcal{M}_k) = k\hat{q} + \sum_{i=1}^k \log(|\hat{\theta}_i|), \quad (2.2)$$

which is seen to depend on the amount of data. The criterion (2.1) may be further minimized over the number of the parameters  $k$  to get the optimal model as well as its complexity in the larger class  $\mathcal{M} = \bigcup_k \mathcal{M}_k$ .

By expanding (2.1) into Taylor's series about the minimizing parameter value, the optimal precision  $\hat{\delta}$  is seen to behave asymptotically as  $1/\sqrt{n}$ , which gives the optimal asymptotic code length approximately as

$$L(x^n | \mathcal{M}_k) = -\log P(x^n | \hat{\theta}) + \frac{k}{2} \log n, \quad (2.3)$$

derived in Rissanen (1978) and by different arguments in Schwarz (1978).

The criterion (2.1) may be interpreted as the Bayesian posterior maximization principle. Indeed, let  $L(\theta, q)$  denote the prefix code length needed to describe the parameters, truncated to the precision  $\delta = 2^{-q}$ . Then  $\pi(\theta^\delta) = 2^{-L(\theta, q)}$  defines a prior for the truncated parameters, and (2.1) is equivalent to maximizing the posterior probability  $P(\theta, q | x^n, \mathcal{M}_k)$  over  $\theta$  and  $q$ . However, this does not mean that the maximum posterior principle is equivalent even with this particular application of the MDL principle. The reason is that the criterion (2.1) is only an approximation to the shortest code length, even if ignore the approximations made in getting  $L(\theta, q)$  which could be removed by starting with a prior  $\pi(\theta^\delta)$  and taking  $L(\theta, q) = -\log \pi(\theta^\delta)$ . The MDL principle calls for the shortest code length for the data  $x^n$ , only, given the class  $\mathcal{M}_k$ , for which (2.1) gives an upper bound. The reason why this is so is that with each parameter value  $\theta$  we can encode all the data, which means redundancy. To remove it, let  $X_\theta$  be the set of all strings of length  $n$  which

the maximum likelihood estimator  $\hat{\theta}(x^n)$  maps to the value  $\theta$ . The sets  $X_\theta$  are disjoint, and the data string has the probability  $P(x^n|\hat{\theta}(x^n))\pi(\hat{\theta}(x^n))/P(X_{\hat{\theta}(x^n)})$ , which with optimal truncation on the parameters gives a shorter code length than (2.1).

## 2.2. Predictive Coding

Suppose we do the coding sequentially as follows: First, order the data set in any manner, unless already done, say as  $x_1 \leq x_2 \leq \dots \leq x_t \leq \dots \leq x_n$ . Then, subdivide the sequence into consecutive blocks of length  $d$ , except possibly the last, the parameter  $d$  to be optimized. To start the procedure, encode the numbers  $x_1, \dots, x_d$  in the first segment any way agreed with the decoder, say by adjoining to the model class a special distribution  $P(x^n|\lambda)$ , where  $\lambda$  represents the empty parameter. Then, recursively, let  $\hat{\theta}(x^{id})$  denote a suitable estimate, often the maximum likelihood one, determined from the first  $i$  segments, and encode the numbers  $x_t$  in the next, the  $i+1$ st segment, with help of the conditional distribution  $P(x_{t+1}|x^t, \hat{\theta}(x^{id}))$ , which can be calculated from the members of the model class. Indeed,  $P(x_{t+1}|x^t, \theta) = P(x^{t+1}|\theta)/P(x^t|\theta)$ . The optimal code length for the data is then to a good approximation given by

$$PMDL(x^n|\mathcal{M}_k) = \min_d \left\{ - \sum_{i \geq 0}^{\min\{(i+1)d-1, n-1\}} \sum_{t=id} \log P(x_{t+1}|x^t, \hat{\theta}(x^{id})) + \log d \right\}, \quad (2.4)$$

where  $\hat{\theta}(x^0) = \lambda$ . Notice that in this predictive code length criterion there is no need to explicitly tell the decoder any parameter values, because they are calculated recursively by an algorithm assumed to be known to him. Neither is there any particular precision needed since the parameters may be calculated to the machine precision.

This model selection criterion does not even need a code length interpretation for its justification, because instead of the code lengths  $-\log P(x_{t+1}|x^t, \hat{\theta}(x^{id}))$  we could just as well have used some prediction error, such as the squared distance

$(x_{t+1} - \hat{x}_{t+1}(\hat{\theta}(x^t)))^2$ , resulting from use of a parametric predictor. This in fact was the way the principle was discovered independently in Dawid (1984), and it was called the 'prequential' principle. The predictive coding is generally very efficient. It is really based upon the sensible expectation that the future behaves as the past. Indeed, if that fails so does everything else! On the negative side, in some applications the ordering requirement of the data imposes a restriction. Again, an arbitrary ordering, in particular for small samples, does affect the criterion. Paradoxically, in some cases, where the prediction error measure is so weak that prediction can be done without knowledge of the entire model, minimization of the predictive criterion does not lead to optimal prediction! An example is the loss function for discrete data, say for binary strings, where a mistake incurs a unit penalty while a correct prediction incurs none. An optimal prediction can be obtained by using the code length criterion to find the optimal model and then using it as the predictor; for an analysis, see Weinberger et al (1993).

### 2.3. Mixture Coding

For the third and the final coding technique we discuss it is necessary to complement the model class with a distribution  $\pi(\theta)$ , which traditionally is called a 'prior'. For us it is just an additional part of the model class which we shall take advantage of to shorten the code length for the data. It actually can be computed from the likelihood function and certainly need not be interpreted as representing prior knowledge. For the present purposes, however, we take it as given. With the so enlarged model class  $\mathcal{M}_k = \{P(x^n|\theta), \pi\}$  define

$$I(x^n|\mathcal{M}_k) = -\log P(x^n|\mathcal{M}_k), \quad (2.5)$$

where

$$P(x^n|\mathcal{M}_k) = \int P(x^n|\theta)\pi(\theta)d\theta. \quad (2.6)$$

A code designed with the code lengths (2.5) is very efficient, for one can show, Rissanen (1987), that (2.5) is strictly smaller than (2.1) for large enough  $n$ , which is



a reflection of the redundancy discussed in Subsection 2.1. However, the integration requirement restricts the applications of the criterion (2.5) to the few but important special distributions where the integral can be evaluated. These are the distributions with the so-called conjugate priors, and they include the gaussian, the multinomial, and the Wishart distributions. The conjugate priors have typically additional so-called nuisance parameters, which can be determined by minimizing the sum of (2.5) and the code length needed to encode these parameters, again truncated optimally. It is curious that the distribution (2.6) has been well-known to the Bayesians, who call it the 'predictive' distribution. However, it was apparently not applied as a model selection criterion until in Rissanen (1987), where the current code length interpretation was given, and, in fact, where (2.5) was defined to be the stochastic complexity.

In a recent talk, given in a Machine Learning Workshop at the University of Pennsylvania, Prof. Breiman described among other things predictors, defined by a convex linear combination of a family of other predictors. He had found empirically that such mixture predictors generally perform very well, better than anyone of the component predictors. We show now that the goodness of the mixture predictor is a simple consequence of the above stated fact that (2.5) is strictly smaller than (2.1) for a finite or countable mixture, or to put it the other way, that the mixture probability or density is strictly larger than any of the terms in the sum. To be specific, consider the regression problem, where we construct a model of the data  $(y^n, \mathbf{x}^n) = (y_1, \mathbf{x}_1), \dots, (y_n, \mathbf{x}_n)$ . Here,  $\mathbf{x}_i$  is a high dimensional vector of the regressor variables  $x_{1i}, x_{2i}, \dots, x_{Ki}$  and  $y_i$  is the response variable. Next, consider the mixture density model

$$f(y^n|\mathbf{x}^n) = \sum_i c_i f(y^n|\mathbf{x}^n, i),$$

where the coefficients  $c_i$  are positive with sum unity, and  $f(y^n|\mathbf{x}^n, i)$  denotes the  $i$ th model, obtained by putting  $f(y^n|\mathbf{x}^n, i) = \prod_t f(y_t|\mathbf{x}_t, i)$ . Each factor is a model obtained, for instance, by taking  $f(y_t|\mathbf{x}_t, i)$  as normal with some variance  $\sigma_i^2$  and the mean given by some predictor  $\phi_i(\mathbf{x}_t)$  of  $y_t$ . It is clear that  $f(y^n|\mathbf{x}^n)$  is strictly larger

than any term  $c_i f(y^n | \mathbf{x}^n, i)$ , including the largest. Since in any successful mixture the coefficients of the largest densities certainly must not be the smallest, the concentration of the mixture probability mass is at or near the observed sample. In fact, in the regression case it can be shown that the mixture density approaches the data generating density, when one is assumed to exist, at the fastest possible rate. Hence, it follows that its conditional mean gives a predictor  $E[Y^n | \mathbf{X}^n] = \sum c_i E[Y^n | \mathbf{X}^n, i]$ , which in effect is unbeatable. Since we tacitly assumed independence we also have

$$\phi(\mathbf{x}_t) \equiv E[Y_t | \mathbf{X}_t] = \sum c_i \phi_i(\mathbf{x}_t),$$

which explains the empirical findings of Breiman.

### 3. Universal Modeling

Traditional estimation theory goes as follows: Fit a given number of parameters to a sample  $x_1, \dots, x_n$  by an estimation procedure, called *estimator*, to give the estimated parameters  $\hat{\theta}(x^n)$  and the model  $P(y | \hat{\theta}(x^n))$ . This model, then, may be applied to new data  $y$  of any size. For analysis purposes one often assumes the new samples to arrive independently from the old, and the behavior of the estimated model, such as the resulting mean prediction error, may be studied. Similarly, the variance of the estimated parameters may be analyzed. A quite extensive theory of this type exists.

There is, however, another way to look at the estimation problem, one which opens up a different vista. For this, we first add one further requirement to the models in the class of interest, namely, that each defines a random process. This is done by imposing for each model the condition

$$\sum_x P(x_1, \dots, x_t, x | \theta) = P(x_1, \dots, x_t | \theta) \quad (3.1)$$

for all  $t$ . Here we used the same letter  $P$  to denote a distribution for any number of arguments. Such a condition is clearly necessary for the models to be any good

in describing the behavior of future data that we on the whole can learn. Now, an estimated model  $P(y|x^n, \hat{\theta}(x^n))$  from the existing data  $x^n$  should be used only to infer the behavior of the very next data item  $y$ , labeled  $x_{t+1}$ . This is because once this item has been seen, we should clearly take advantage of the additional information it provides to obtain a better model, and so on. But since we know that condition (3.1) is necessary to infer the behavior of  $x_{t+1}$  from the past  $x^n$  our modeling procedure actually should describe a random process  $\hat{P}(x^n)$ . We deliberately omit including parameter estimates  $\hat{\theta}(x^n)$  in this process, because it may not use any! Clearly, the procedure that describes the random process should be independent of the data, which means that it should provide a good model of whatever machinery generates the data. In particular, for the purposes of analysis, where we often assume that the data are samples from some process, the modeled random process should mimic the actual data generating process, whatever it is. In other words, the modeled process should have the *universality* property that it is capable of imitating the behavior of any data generating process in the considered family. Clearly, for that to be possible the modeled process itself cannot belong to the family. The three central questions that then arise are whether such universal processes exist at all, and if yes how well the imitation can take place, perhaps, in an asymptotic sense, and finally whether we actually can construct an optimal universal process. In a very real sense the construction of an optimal universal process settles the modeling problem: Use the process for all the modeling tasks such as prediction, decision, or control, just as if it were the 'true' data generating process.

A number of results have been proved during the recent years which shed light to these questions, Rissanen (1984), (1986a), (1986b), Hannan et al (1989), Gerencser (1989), Hemerly and Davis (1989), Yu (1990). An explicit and efficient construction of a universal process for the class of Markov chains of finite order can be done with the algorithm Context, Rissanen (1983b), (1993), Furlan (1989). A related one was proved to be asymptotically optimal in Weinberger et al (1993). As an illustration, we give one such result.

**Theorem 2.1.** Consider a class of models  $\mathcal{M} = \bigcup_k \mathcal{M}_k$ , where  $\mathcal{M}_k = \{P(x^n|\theta)\}$ , the parameter  $\theta$  ranging over a compact subset  $\Gamma^k$  of  $R^k$  with nonempty interior. Let  $P(x^n|\theta)$  satisfy the marginality condition (3.1) and be smooth enough to admit an estimator  $\hat{\theta}(x^n)$  which satisfies the central limit theorem. Then (i) for any distribution  $P(x^n)$  satisfying (3.1)

$$\lim_{n \rightarrow \infty} \frac{E_\theta \log(P(x^n)/P(x^n|\theta))}{(k/2) \log n} \geq 1 \quad (3.2)$$

for all  $k$  and  $\theta$  in  $\Gamma^k$ , except in a null set. The expectation is taken with respect to the distribution  $P(x^n|\theta)$ . Moreover, (ii) there exists a distribution  $P^*(x^n)$ , satisfying (3.1), for which (3.2) holds with equality.

To put the theorem slightly differently and less formally, there is a universal random process  $P^*(x^n)$  such that the ideal code length it defines satisfies

$$-\frac{1}{n} E_\theta \log P^*(x^n) = -\frac{1}{n} E_\theta \log P(x^n|\theta) + \frac{k \log n}{2n} + o\left(\frac{\log n}{n}\right)$$

no matter which process  $P(x^n|\theta)$  generates the sequences. Moreover, for all intents and purposes no distribution can do better. Finally, one may even drop the expectation operations, and the result holds essentially for all sequences generated by the data generating process. When  $k = 0$  we get in essence Shannon's theorem. Finally, a corollary of this theorem provides a tight lower bound for the mean square prediction error, Rissanen (1984).

We conclude this section by describing briefly a modification of the algorithm Context, which implements a universal process for time series of the type just discussed. In broad terms the idea is to construct recursively in  $t$ , as the data  $x_t$  are obtained, a state space, its partition, and a process of the type

$$x_t = F(x_{t-1}, \dots, x_{t-k}) + e_t, \quad (3.3)$$

where  $k$  may depend on the past string  $x^{t-1}$  and is to be optimized. Begin by truncating each observation  $x_t$ , for simplicity, to a binary number  $\bar{x}_t$ , called a *symbol*. The algorithm grows a binary tree by the rules:

1. Start with the one-node tree  $T_0$  with its two symbol occurrence counts initialized to 0.
2. Recursively, having constructed the tree  $T_t$  from  $\bar{x}^t$ , climb the tree along the path  $\bar{x}_t, \bar{x}_{t-1}, \dots$  into the past. For each node  $s$  visited, update the count of the symbol  $\bar{x}_{t+1}$  by one:

$$c_{t+1}(\bar{x}_{t+1}|s) = c_t(\bar{x}_{t+1}|s) + 1.$$

3. If the last updated count for the node, say  $\bar{x}_t, \bar{x}_{t-1}, \dots, \bar{x}_{t-k}$ , becomes at least 2, extend the tree by the node  $\bar{x}_t, \bar{x}_{t-1}, \dots, \bar{x}_{t-k}, \bar{x}_{t-k-1}$ , and initialize its symbol counts to zero, except the count for the symbol  $\bar{x}_{t+1}$ , which is set to 1. This gives the tree  $T_{t+1}$ .

While growing the tree the algorithm also fits an AR model to the past occurrences of the *original* full precision numbers at each node  $s$ , to give the predictors

$$\hat{x}_{t+1} = a_0(t, s) + a_1(t, s)x_t + \dots + a_r(t, s)x_{t-r+1}, \quad (3.4)$$

where the order  $r$  is optimized by the MDL principle. In particular, it is not restricted to be less than or equal to the depth of the node. The coefficients  $a_i(t, s)$  will be functions of the past occurrences of observations, say  $C_t(s)$ , in the node  $s$ , which clearly defines an equivalence class. With these we may then calculate the sum of the 'honest' prediction errors  $L_t(s) = \sum_{\tau \in C_t(s)} (x_\tau - \hat{x}_\tau(s))^2$  from the occurrences of the past symbols in this node. These are used to find the optimal node  $s_t^*$  for the symbol  $x_{t+1}$  by minimization over the nodes along the path  $x_t, x_{t-1}, \dots$ . The predictors (3.4) define a process of type (3.3) if we put

$$x_{t+1} = \hat{x}_{t+1}(s_t) + e_{t+1}. \quad (3.5)$$

Since the depth of the optimal node may well be smaller than the order of the AR model in this node, the resulting representation is smoother than piecewise linear, even though only linear fits are being calculated.

#### 4. EXAMPLES

We illustrate the MDL principle with two simple examples.

**Example 1.** We applied the algorithm Context to the data of length 500 generated by the following nonlinear AR system,

$$x(t+1) = \alpha x(t-1)|x(t)|^{1/2} + \beta x(t)|x(t-2)|^\gamma + e(t), \quad t = 0, \dots, n-1, \quad (4.1)$$

where  $e(t)$  was obtained as follows: First, a gaussian sample sequence with zero mean and unit variance was generated. Then the positive outcomes were multiplied by three, and the sequence was centered by subtracting the sample arithmetic mean from all the numbers. The intent was to have something else than symmetric gaussian noise. The three parameters had the values  $\alpha = -.5$ ,  $\beta = .15$ , and  $\gamma = 1.55$ , which brought the system near its stability boundary. The three initial values needed were  $x(0) = x(-1) = x(-2) = 0$ . The sample variance of the noise and the output data was 4.25 and 6.18, respectively.

We first fitted linear AR models. The best PMDL determined order was 2 with the predictive variance  $\hat{\sigma}_{AR}^2 = 5.72$ , which at the same time served as the value of the predictive MDL criterion. Algorithm Context, in turn, gave the substantially smaller value  $\hat{\sigma}_{CX}^2 = 5.30$ . Hence, by the MDL principle we should prefer the nonlinear model delivered by Algorithm Context. To see whether the principle is reliable, we then took the best linear AR model and applied it to predict the values in a new sample of length 500 generated by the same system (4.1) using the corresponding linear predictor. The result was the per symbol squared prediction error  $PE_{AR} = 5.39$ . We then predicted the same sample with the best model found by Algorithm Context with the substantially smaller result  $PE_{CX} = 4.85$ , just as could have been anticipated by the values of the predictive criterion calculated from the first sample.

**Example 2.** Table 1 shows a two-way contingency table, where the entries  $n_{ij}$  indicate the observed cell occurrences of the pair  $(x_i, y_j)$  of attribute values in a sequence  $(x, y)$  of length  $n$ . For example, in Kendall and Stuart (1961, page 552) the influence of the feeding habits of children to the nature of their teeth was studied,

where  $(x_1, y_1)$  means breast feeding and normal teeth,  $(x_1, y_2)$  breast feeding and abnormal teeth,  $(x_2, y_1)$  bottle feeding and normal teeth, and  $(x_2, y_2)$  bottle feeding and abnormal teeth.

$(x, y)$	$y_1$	$y_2$	Totals
$x_1$	$n_{11}$	$n_{12}$	$n_{1.}$
$x_2$	$n_{21}$	$n_{22}$	$n_{2.}$
Totals	$n_{.1}$	$n_{.2}$	$n$

**Table 1. A two-way contingency table**

The most frequently tested hypotheses in such a table are whether the two attributes are independent or not. Both hypotheses are modeled by a distribution in which the cell occurrences take place with probabilities  $p_{ij}$ , which act as parameters, and the cell occurrences are independent so that the probability of the string is the product of the cell probabilities. The null-hypothesis is represented by the model class  $\mathcal{M}_0$ , which states that the four cell probabilities satisfy the independence condition of being given by the product of the marginal probabilities  $p_{ij} = p_i \cdot p_j$ . Each parametric distribution in the model class is then the product of two Bernoulli distributions, one for the columns and the other for the rows in the table. Taking the uniform prior for each we get from (2.5)

$$I(x, y | \mathcal{M}_0) = \log \left[ \binom{n}{n_{1.}} \binom{n}{n_{.1}} \right] + 2 \log \binom{n+1}{n}. \quad (4.2)$$

The alternative hypothesis is represented by the model class  $\mathcal{M}_1$ , defined by the distributions with three free parameters  $p_{ij}$ ,  $i, j = 1, 2$ , which satisfy only the constraints that they are non-negative and add up to unity. Again with the uniform prior in the range of the free parameters, we get

$$I(x, y | \mathcal{M}_1) = \log \binom{n}{n_{ij}} + \log \binom{n+3}{n}. \quad (4.3)$$

By the MDL principle the winning hypothesis; ie, model class, is the one which gives the shorter code length for the data string  $(x, y)$ . Because of the type of model classes chosen, the probabilities assigned to the string by the models in each class depend only on the occurrence counts in Table 1, and we can readily calculate the code length difference

$$I(x, y | \mathcal{M}_0) - I(x, y | \mathcal{M}_1) = \log \frac{n! n_{11}! n_{12}! n_{21}! n_{22}!}{n_1! n_2! n_{\cdot 1}! n_{\cdot 2}!} - \log \frac{(n+2)(n+3)}{3!(n+1)},$$

which serves as a universal test statistic  $T(x, y)$ . With the numerical values  $n_{11} = 4$ ,  $n_{12} = 16$ ,  $n_{21} = 1$ , and  $n_{22} = 21$  we get  $T(x, y) = 0.056$ , and the independence hypothesis is narrowly rejected. We see that our test is like the traditional test, and the non-negative random variable defined by the first term in  $T(x, y)$  is matched with the positive threshold, defined by the second term. The ratio in the first term is a uniform most powerful unbiased test statistic, Kendall and Stuart (1961, Section 34.24), which, evidently, is equivalent with our test statistic. This is not an accident; it holds whenever such test statistics exist, which clearly provides a powerful support to the reasonableness of our utterly simple testing procedure.

Due to the smallness of the test statistic we would expect our confidence in rejecting the null-hypothesis to be low. How to form a realistic measure of this confidence? We have argued in Rissanen (1989) that the very best way to assess the confidence would be to repeatedly gather samples like Table 1, generated by the same 'physical machinery' with which this table was obtained; ie, to have more data, and then do the test again and again. The distribution of the test statistic would give us the probability of our having made a mistake. However, in most cases this is not possible, and we would like to do the next best thing, which is to model the physical machinery and sample that. Clearly, the goodness of the results depends critically on our ability to construct a good model of the physical machinery so that the new data were statistically similar to the actually observed sample. But the entire purpose of the MDL principle is to get best models of data, and in this instance we must take the optimal model in the non-independent model class  $\mathcal{M}_1$  as the one with which to generate the new data. We generated



200 repetitions of samples of size 42 with the model defined by the cell probabilities  $\hat{p}_{11} = 4/42$ ,  $\hat{p}_{12} = 16/42$ ,  $\hat{p}_{21} = 1/42$ , and  $\hat{p}_{22} = 21/42$ . The null hypothesis was accepted with probability 0.46, which, indeed is close to a toss-up, indicating very small confidence on our test result.

The described procedure is clearly similar to Effron's bootstrap, except for our requirement that the new data be generated with the best model we can find. Indeed, without this requirement we see no justification for the technique. In fact, even so all we can assess is the uncertainty due to sample fluctuations relative to the chosen model. The other source of uncertainty, that due to the lack of our model not being perfect or even optimal, will always remain beyond reach. As we discussed in Section 1, this, indeed, is the ultimate uncertainty in all model building, because the issue of finding the optimal model from data is undecidable. We should add that in all but simple cases this second source of uncertainty is the dominant one, and the accurately calculated confidence intervals provide a false sense of confidence.

## Appendix

Let  $A$  denote a finite or countable set called an *alphabet*. Its elements are called *symbols*, which we frequently in the case of a finite alphabet identify with the first  $d + 1$  integers  $0, 1, \dots, d$ . Write  $A^n$  for the set of all strings of length  $n$  and  $A^* = \bigcup_{n=0}^{\infty} A^n$  for their union. For convenience, the first power  $A^0$  consists of the empty string, written as  $\lambda$ . In information theory a finite string  $x = a_1, \dots, a_n \in A^*$  of symbols is called a *message*, but we prefer the name *data string* or sequence.

A *code*  $C$  is a one-to-one map from  $A^*$  into  $B^*$ , the set of all finite binary strings. Nothing essential is lost by restricting the code alphabet to be binary, which for our purposes is all that is needed. A simple example is a code defined for the three-symbol alphabet  $A = \{a, b, c\}$  as follows:  $C(a) = 0$ ,  $C(b) = 10$ ,  $C(c) = 11$ . This is extended to strings by replacing the symbols by their corresponding codewords thus:  $C(aabac) = 0010011$ . Notice that with this particular code you will be able to

decode any binary string without commas separating the successive codewords. This is possible because the code defines a binary subtree where the leaves correspond to the three codewords, and no extension of a codeword can define another codeword. When each codeword is a leaf, the code is called a *prefix* code, or it is said to have the *prefix* property. Such a property is not only desirable for the sake of easy and 'instantaneous' decoding, but it implies the fundamental *Kraft inequality* for the code lengths

$$\sum_{x \in A} 2^{-L(x)} \leq 1 \quad (\text{A.1})$$

for a finite or even countable alphabet, where the code length  $L(x)$  is the number of digits in the codeword  $C(x)$ . The equality holds if and only if the tree, defined by the codewords, is complete in the sense that there is no leaf which is not a codeword, in which case the code is called a *complete prefix* code. An easy proof of (A.1) is done by induction on the number of leaves.

Suppose next that we are given  $d+1$  positive integers  $n_0, \dots, n_d$  satisfying the Kraft-inequality

$$\sum_{i=0}^d 2^{-n_i} \leq 1,$$

and we ask whether it is possible to construct a prefix code for the alphabet  $\{0, \dots, d\}$  with lengths defined by these integers. The answer, of course, is yes. All we need to do is to sort the integers by increasing size, and construct the code tree as follows: Assign to the first codeword the left-most leaf  $0 \dots 0$  of path length given by the smallest integer. Continue by assigning to the next codeword the next left-most available leaf of length defined by the second smallest integer (which, of course, may be the same as the smallest), and so on. The Kraft-inequality guarantees that there always will be enough nodes for the codewords, regardless of the alphabet size. Clearly, this is not the only code with the given lengths.

We see in (A.1) that a prefix code defines a distribution via  $P(x) = K 2^{-L(x)}$ , where  $K = 1 / \sum_y 2^{-L(y)}$  is the normalizing coefficient needed in case the code is

not complete. Conversely, if we have a distribution defined on a finite or countable set, we can construct a prefix code such that its codeword lengths coincide with the integers  $\lceil -\log P(x) \rceil$ , where  $\lceil y \rceil$  denotes the smallest integer upper bound to the number  $y$ . Hence, to within the normalization a prefix code and a distribution are equivalent. Since any finitely describable object; ie, an object that can be associated with a finite string over a finite alphabet, can surely be encoded with a codeword of a prefix code, we can also talk about the so-defined probability of the object. Further, since it does not make any difference whether we encode 'random' data or 'nonrandom' parameters, we have a uniform interpretation of probabilities in terms of the code lengths, which is in contrast with the Bayesian philosophy, where the interpretation of probabilities for the parameters poses grave difficulties calling for 'subjective' or other nonscientific means. As a practical matter, it is sometimes far easier to contemplate concrete codes for objects, frequently parameters about which no repeated data are available, and calculate their code lengths than to select more or less arbitrary distributions as 'priors' for them. For example, we may wish to talk about polygons on a plane. It is easy to visualize how to encode each by encoding the position of its nodes. Compare this with the task of selecting a 'prior' distribution for the set of all polygons!

We next establish a link between the code length and entropy by proving the first fundamental theorem in information theory, usually credited to Shannon but also sometimes referred to as Gibbs' inequality.

**Theorem A1.** Let  $S$  be a finite or countable set, and let  $P$  and  $Q$  be two distributions on  $S$ . Then

$$(i) \quad -\sum_{x \in S} P(x) \log Q(x) \geq -\sum_{x \in S} P(x) \log P(x) \equiv H(X).$$

Moreover, the equality holds if and only if

$$(ii) \quad Q(x) = P(x)$$

for every  $x$ . Here,  $0 \log 0 = 0$ .

**Proof.** By Jensen's inequality,

$$\sum_{x \in S} P(x) \log \frac{Q(x)}{P(x)} \leq \log \sum_{x \in S} P(x) \frac{Q(x)}{P(x)} = 0,$$

the equality (ii) holding as claimed, because the logarithm is strictly concave.

Since each prefix code length function  $L(x)$  defines the distribution  $Q(x) = 2^{-L(x)} / \sum_y 2^{-L(y)}$ , Theorem 1 gives

$$\sum_{x \in S} P(x) L(x) \geq H(X) + \log \sum_y 2^{-L(y)} \geq H(X),$$

so that the entropy is a lower bound for any mean prefix code length, which by (ii) is reached only when the code lengths reflect the data generating distribution. Often the set  $S$  is taken as the set  $A^n$  of all strings of some length  $n$  over an alphabet  $A$ , which is either finite or countable. Moreover, the probability function  $P_n(x^n)$  is in addition required to satisfy the compatibility condition

$$\sum_{z \in A} P_{n+1}(x^n z) = P_n(x^n) \tag{A.2}$$

for all strings  $x^n = x_1, \dots, x_n$ , where  $x^n z = x_1, \dots, x_n, z$ . Such a family of distributions  $\{P_n(x^n)\}$ , also written more simply as a function  $P(x)$  on  $A^* = \bigcup_n A^n$ , defines a random process or an *information source*. Indeed, such a condition is necessary for a model to be useful, because it permits the definition of the conditional probability  $P(z|x^n) = P_{n+1}(x^n z)/P_n(x^n)$ , and hence it provides a link from the past into the future.

Notice that the code length  $-\log P(x)$  given by (ii) of Theorem A1 is optimal only in the sense of the mean, rather than for each individual outcome  $x$ . However, for large  $n$  the set  $S = A^n$  is very large, and just as in long series of flips of a fair coin the ratio of the heads to the total number of throws is close to  $1/2$  in virtually all of them, the overwhelming majority of the strings  $x^n$  generated by the source

are such that  $-\log P(x^n) \approx nH(X)$ . The reason for this is fundamentally the fact about binary trees that the overwhelming majority of the nodes lie near the leaves. Therefore, if we design a code such that it assigns the length  $-\log P(x^n)$  to the particular data string  $x^n$ , we know that it takes a near miracle to find a shorter code for this string, or to put it the other way, the string  $x^n$ , generated by the given source, would have to be an exceptional one for us to be able to encode it with a shorter code length than

$$L(x^n) = -\log P(x^n),$$

which justifiably is also called the *ideal* code length. The word 'ideal' also frees us from the petty requirement that a code length must be an integer.

We conclude this appendix with a brief discussion of how to encode objects, which can be represented as integers, in a prefix manner when no distribution is given for them. Such a code, then, by the Kraft inequality defines a distribution with a certain asymptotically optimal universality property, Rissanen (1983a). First, a binary representation of the integer  $n$  has about  $\log n$  digits. But if such a binary string was followed by other binary symbols we would not be able to read off that integer, because we would not know the length of the binary representation of  $n$ . To remedy the situation we could attach a preamble telling the required length, which requires about  $\log \log n$  binary digits. Iterating this, we attach further preambles telling the length of the length etc until the shortest preamble is reached. We don't really need the exact code, described in detail in Elias (1975), but only the fact that the total number of digits required to encode  $n$  in a prefix manner is about  $\log^* n = \log n + \log \log n + \dots$ , where the sum includes only positive terms. This induces a distribution  $P^*(n) = c2^{-\log^* n}$ , where the normalizing constant is between 2 and 4. Depending on the size of the integers to be encoded, we may approximate  $\log^* n$  by the lower bound  $\log n$ , which coincides with the nonprefix code length resulting from Jeffreys' improper prior  $1/n$ , or by the upper bound  $\log n + 2 \log \log n$ , which of course is a prefix code length.

## REFERENCES

- Chaitin, G.J. (1975), 'A Theory of Program Size Formally Identical to Information Theory', *J. Assoc. Comput. Machines*, **22**, 329-340
- Dawid, A.P. (1984), 'Present Position and Potential Developments: Some Personal Views, Statistical Theory, The Prequential Approach', *J. Royal Stat. Soc. A*, Vol. **147**, Part 2, 278-292
- Dengler, J. (1990), 'Estimation of Discontinuous Displacement Vector Fields with the Minimum Description Length Criterion', MIT A.I. Lab., Memo No. 1265
- Elias, P. (1975), 'Universal Codeword Sets and Representations of the Integers', *IEEE Trans. Inf. Thy*, Vol **IT-21**, no. 2, 194-203
- Furlan, G. (1989), *Contribution a l'Etude et au Developpement d'Algorithmes de Traitement du Signal en Compression de Donnees et d'Images*, PhD Dissertation, l'Université de Nice, Sophia Antipolis, France (in French)
- Gao, Q. and Li, M. (1989), 'An Application of Minimum Description Length Principle to online Recognition of Handprinted Alphanumerals', *Proc. of 11th International Joint Conference on Artificial Intelligence*, Detroit, Michigan, 843-848, Kaufmann Publ.
- Gerencse'r, L. (1989), 'On a Class of Mixing Processes', *Stochastics*, Vol. **26**, 165-191
- Hannan, E.J., McDougall, A.J. and Poskitt, D.S. (1989), 'Recursive Estimation of Autoregressions', *J. Royal Statist. Soc., Ser. B*, **51**, No. 2, 217-233
- Hannan E.J. and Rissanen J. (1988), 'The Width of a Spectral Window', *A Celebration of Applied Probability*, ed. J. Gani, 301-307
- Hemerly, E.M. and Davis, M.H.A. (1989), 'Strong Consistency of the PLS Criterion for Order Determination of Autoregressive Processes', *Annals of Statistics*, Vol. **17**, No. 2, 941-946
- Kendall, M.G. and Stuart, A. (1961), *The Advanced Theory of Statistics*, Vol. **2**, Hafner Publishing Co., New York.

- Kolmogorov, A.N. (1965), 'Three Approaches to the Quantitative Definition of Information', *Problems of Information Transmission* 5, No. 1, 1-7
- Leclerc, Y.G. (1989), *The Logical Structure of Image Discontinuities*, PhD Dissertation, Dept. of EE, McGill University
- Li, M. and Vitanyi, P.M.B. (1992), 'Inductive Reasoning and Kolmogorov Complexity', *J. of Computer and System Sciences*, Vol. 44, No. 2, 343-384
- Quinlan, J.R. and Rivest, R.L. (1989), 'Inferring Decision Trees Using Minimum Description Length Principle', *Information and Computation*, 80, 227-248
- Rao R.B., Lu S. C-Y., and Stepp R.E. (1991), 'Knowledge-Based Equation Discovery in Engineering Domains', *Proc. of the eighth International Workshop on Machine Learning*, pp 630- 634
- Rissanen, J. (1978), 'Modeling by shortest data description', *Automatica*, Vol. 14, pp. 465-471
- Rissanen, J. (1983a), 'A Universal Prior for Integers and Estimation by Minimum Description Length', *Annals of Statistics*, Vol. 11, No. 2, 416-431
- Rissanen, J. (1983b), 'A Universal Data Compression System' *IEEE Trans. Information Theory*, Vol. IT-29, nr 5, pp 656-664, 1983
- Rissanen, J. (1984), 'Universal Coding, Information, Prediction, and Estimation', *IEEE Trans. Inf. Theory*, Vol. IT-30, Nr. 4, 629-636
- Rissanen, J. (1986a), 'Stochastic Complexity and Modeling', *Annals of Statistics*, Vol 14, 1080-1100
- Rissanen, J. (1986b), 'A Predictive Least Squares Principle', *IMA Journal of Mathematical Control and Information*, Vol. 3, Nos 2-3, 211-222
- Rissanen, J. (1987), 'Stochastic Complexity', *The Journal of the Royal Statistical Society, Series B*, Vol. 49, No. 3, 223-239, and 252-265, (with discussions)
- Rissanen, J. (1989), *Stochastic Complexity in Statistical Inquiry*, World Scientific Publ. Co., New Jersey, (175 pages)

- Rissanen, J. (1993), 'Noise Separation and MDL Modeling of Chaotic Systems', Proc. of the NATO Summer School 'From Statistical Physics to Statistical Inference and Back', Cargese, France, 1992.
- Rissanen, J. and Ristad, E. (1992), 'Unsupervised Classification with Stochastic Complexity', Proc. of the US/Japan Conference on the Frontiers of Statistical Modeling: An Informational Approach, U. of Tennessee, May 1992
- Sheinvald, J., Dom, B., Niblack, W., Banerjee, S. (1992), 'Combining Edge Pixels into Parameterized Curve Segments using the MDL Principle and the Hough Transform', a chapter in the book *Advances in Image Analysis*, ed. Y. Mahdavih and R.C. Gonzales, Publ. SPIE
- Schwarz, G. (1978), 'Estimating the Dimension of a Model', *Annals of Statistics* 6, 461-464
- Solomonoff, R.J. (1978), 'A Formal Theory of Inductive Inference' Parts 1 and 2, *Information and Control*, 7, 1-22, 224-254
- Wax, M. and Ziskind, I. (1988), 'Detection of Fully Correlated Signals by the MDL Principle' Proc. of ICASSP 88, New York, 2777-2780
- Weinberger, M.J., Rissanen, J., Feder, M. (1992), 'A Universal Finite Memory Source', submitted to *IEEE Trans. Inf. Theory*
- Yu, B. (1990), *Some Results on Empirical Processes and Stochastic Complexity*, PhD Thesis, Dept. Statistics, UC Berkeley.



# Approximate One-Sided Tolerance Limits for a Mixed Model With a Nested Random Effect

Mark G. Vangel  
U.S. Army Research Laboratory  
Materials Directorate  
AMSRL-MA-DB  
Arsenal St., Watertown MA 02172-0001

## Abstract

Consider the mixed model

$$Y_{ijk} = x_i^T \theta + b_{ij} + e_{ijk},$$

where  $i = 1, \dots, l$ ,  $j = 1, \dots, m_i$ ,  $k = 1, \dots, n_{ij}$ , and the mutually independent random variables  $b_{ij} \sim N(0, \sigma_b^2)$  and  $e_{ijk} \sim N(0, \sigma_w^2)$  denote between-batch and within-batch components of variance, respectively. Based on data  $\{Y_{ijk}\}$ , we will show how to determine approximate one-sided confidence limits on any quantile of the population of the random variable

$$U \sim N(w^T \theta, \sigma_b^2 + \sigma_w^2),$$

where  $w$  is an arbitrary known vector.

Lower confidence limits on lower tail quantiles of a population of material strength measurements are routinely used to characterize the strength of a material, particularly in aircraft design. Composite materials typically exhibit considerable between-batch variability, so that the methodology discussed in this article could have important applications. For example, if three batches of five specimens each are tested at each of four temperatures, and it is desired to determine *as a function of temperature* a lower confidence limit on the tenth percentile of the population corresponding to the strength of a specimen chosen at random from a randomly selected batch, then the proposed methodology could be applied.

## 1 Introduction

In structural design, an *allowable stress*, *working stress*, or *design allowable* for a material is the maximum stress at which one can be reasonably certain that failure will not occur. For the design of structures for which weight is not a primary consideration, allowables are typically calculated by dividing a stress level at which failure is known to often occur by a sufficiently large constant called a *safety factor* (e.g., Gere and Timoshenko, 1984, p.29). The structure is then designed so as to ensure that the stresses do not exceed the allowables for the materials.

This approach is too conservative for most aircraft applications, however. Since weight is an important consideration in aircraft design, this industry long ago established two one-sided tolerance limits to supplement the use of safety factors in determining allowables. These tolerance limits are a 95% lower confidence limit on the tenth percentile and a 95% lower confidence limit on the first percentile of the strength distribution of a material. These are referred to as 'B-basis' and 'A-basis' values, respectively (Mil Handbook 5E, 1987; Mil Handbook 17C, 1992).

Composite materials are being used increasingly often in aircraft. These materials can provide the strength and stiffness of metallic components at substantially less weight. Composite material strength can vary from batch to batch, and tolerance limits based on pooled data can be dangerously optimistic. Consequently, procedures for tolerance limits in the presence of between-batch variability are of considerable importance in aircraft design.

The literature on random-effects tolerance limits is largely confined to the one-way balanced ANOVA model (Mee and Owen, 1983; Vangel 1992). Although an understanding of this simple model has been an important first step, it is necessary to make progress toward general methodology which can cope with unbalanced designs and covariates. This article takes a step in this direction, by proposing an approximate method for obtaining one-sided random-effects tolerance limits for an arbitrary mixed model with a nested random effect. The testing of composite materials is expensive, and engineers can usually only obtain a small amount of data for each value of various fixed effects (e.g., three batches of five specimens at each of several temperatures). The usual approach to calculating tolerance limits for such data involves

regression methods (e.g., Owen (1968), pp. 462-463) which ignore the batch effect. In this article we introduce an approach which includes the random batch effect and we apply this method successfully to two real-data examples.

## 2 The Model

Assume that we have data  $\{Y_{ijk}\}$ , where

$$Y_{ijk} = x_i^T \theta + b_{ij} + e_{ijk}, \quad (1)$$

for  $i = 1, \dots, l$ ,  $j = 1, \dots, m_i$ , and  $k = 1, \dots, n_{ij}$ . We will adopt the usual convention of indicating summation over a subscript by a dot, e.g.  $n_i = \sum_j n_{ij}$ . The independent random variables  $\{b_{ij}\}$  and  $\{e_{ijk}\}$ , distributed

$$b_{ij} \sim N(0, \sigma_b^2) \quad (2)$$

and

$$e_{ijk} \sim N(0, \sigma_w^2), \quad (3)$$

model between-batch and within-batch random effects, respectively. The  $r \times 1$  vectors  $\{x_i\}$  are arbitrary; though for convenience we will assume that the  $r \times n_{..}$  matrix  $X$ , which consists of rows  $x_i^T$  each repeated  $n_i$  times, is of rank  $r$ , so that  $X^T X$  is nonsingular.

In terms of this matrix  $X$ , we can write (1) as

$$Y = X\theta + \eta, \quad (4)$$

where

$$\eta \sim N(0, \Sigma), \quad (5)$$

$$\Sigma = \text{diag}(\Sigma_{11}, \dots, \Sigma_{lm_l}), \quad (6)$$

and

$$\Sigma_{ij} = \sigma_w^2 I + \sigma_b^2 J, \quad (7)$$

for  $J$  a  $n_{ij} \times n_{ij}$  matrix of ones. Let  $\hat{\theta}$  be the ordinary least squares estimator of  $\theta$ , so that

$$\hat{Y} = X\hat{\theta} = X(X^T X)^{-1} X^T Y \equiv HY, \quad (8)$$

where  $H$  (the 'hat matrix') is a projection matrix.

### 3 Estimating $\sigma_b^2$ and $\sigma_w^2$

Because the matrix  $X$  consists of  $l$  distinct rows repeated in blocks of  $n_{i.}$  rows each, it can be shown that the matrix  $H$  will be constant along diagonal blocks of size  $n_{i.} \times n_{i.}$ . Since  $H$  has this 'block constant' structure, we can easily determine closed form expressions for  $\text{tr}(H\Sigma)$  and  $\text{tr}[(H\Sigma)^2]$  in terms of the  $h_{i.}$ . This will enable us to calculate the means and variances of certain quadratic forms (see, e.g., Seber (1977), Section 1.4), and to thereby derive estimators of  $\sigma_b^2$  and  $\sigma_w^2$ .

The value of  $H$  in these blocks are the  $l$  distinct 'hat matrix diagonals', available from most least squares regression software, and we will follow convention and denote these as  $\{h_{i.}\}_{i=1}^l$ .

We begin by defining a second model, in which the matrix  $X$  is augmented to an  $(r + m, -l) \times n_{..}$  matrix  $\tilde{X}$  by the addition of  $m, -l$  columns of batch indicators:

$$Y = \tilde{X}\tau + \epsilon, \quad (9)$$

where

$$\epsilon \sim N(0, \sigma_w^2 I), \quad (10)$$

and  $\hat{\tau}$  denotes the least squares estimate of  $\tau$ ,

$$\hat{\tau} = (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T Y. \quad (11)$$

We will use the first two moments of the residual sums of squares from these two models,

$$\text{RSS}_A = (Y - X\hat{\theta})^T (Y - X\hat{\theta}) \quad (12)$$

and

$$\text{RSS}_B = (Y - \tilde{X}\hat{\tau})^T (Y - \tilde{X}\hat{\tau}), \quad (13)$$

in order to construct estimators of the two components of variance. Because of the block diagonal structure of  $\Sigma$  and the 'block constant' structure of  $H$ , it is straightforward to calculate these moments:

$$E(\text{RSS}_A) = \sum_{ij} [(1 - n_{ij}h_{i.})(n_{ij}\sigma_b^2 + \sigma_w^2) + (n_{ij} - 1)\sigma_w^2], \quad (14)$$

$$\text{Var}(\text{RSS}_A) = 2 \sum_{ij} [(1 - n_{ij}h_{i.})(n_{ij}\sigma_b^2 + \sigma_w^2)^2 + (n_{ij} - 1)\sigma_w^4], \quad (15)$$

$$E(\text{RSS}_B) = \sigma_w^2(n_{..} - r - m, + l), \quad (16)$$

$$\text{Var}(\text{RSS}_B) = 2\sigma_w^4(n_{..} - r - m, + l). \quad (17)$$

We note that

$$\text{RSS}_B \sim \sigma_w^2 \chi_{n_{..} - r - m_{..} + l}^2, \quad (18)$$

and we conjecture that

$$\text{RSS}_A \sim \sum_{ij} (n_{ij} \sigma_b^2 + \sigma_w^2) \chi_{1 - n_{ij} h_i}^2 + \sigma_w^2 \chi_{n_{..} - m_{..}}^2. \quad (19)$$

The residual sums of squares  $\text{RSS}_A$  and  $\text{RSS}_B$  are generalizations of the 'total' and 'within' sums of squares in the one-way ANOVA model. However, in this case we do *not* have the usual decomposition of ANOVA sums of squares; in particular  $\text{RSS}_A$  and  $\text{RSS}_B$  are *not* independent. It is possible to construct estimators of the variance components in terms of sums of squares which *are* independent, with only superficial changes to the proposed methodology. We have chosen to use the sums of squares defined above because, when the model is correct, these sums of squares have more precision than the independent sums of squares. However, when the model is wrong,  $\text{RSS}_A$  and  $\text{RSS}_B$  will be biased; so there is an implicit tradeoff between bias and variance involved in the decision of how to estimate the variance components.

We normalize the residual sums of squares (12) and (13), giving the mean squares

$$\text{RMS}_A \equiv \frac{\text{RSS}_A}{\sum_{ij} n_{ij} (1 - h_i)} = \frac{\text{RSS}_A}{n_{..} - r} \quad (20)$$

and

$$\text{RMS}_B \equiv \frac{\text{RSS}_B}{n_{..} - m_{..} - r + l} \sim \sigma_w^2 \frac{\chi_{n_{..} - m_{..} - r + l}^2}{n_{..} - m_{..} - r + l}, \quad (21)$$

since

$$\sum_{ij} n_{ij} (1 - h_i) = n_{..} - \text{tr}(H) = n_{..} - r. \quad (22)$$

Since  $E(\text{RMS}_B) = \sigma_w^2$ ,  $\text{RMS}_B$  provides an unbiased estimator of  $\sigma_w^2$ . Define

$$\text{RMS}_A^* \equiv \frac{\sum_{ij} n_{ij} (1 - h_i)}{\sum_{ij} n_{ij} (1 - n_{ij} h_i)} \text{RMS}_A. \quad (23)$$

When  $\sigma_b = 0$ ,  $E(\text{RMS}_A) = \sigma_w^2$ , and when  $\sigma_w^2 = 0$ ,  $E(\text{RMS}_A^*) = \sigma_b^2$ . When  $\sigma_b^2 = 0$ ,  $\text{RMS}_A$  has (at least) the same first two moments as

$$\text{RMS}_A^0 \sim \sigma_w^2 \chi_{\nu_0}^2 / \nu_0, \quad (24)$$

where

$$\nu_0 = \sum_{ij} n_{ij}(1 - h_i) = n_{..} - r. \quad (25)$$

When  $\sigma_w^2 = 0$ ,  $\text{RMS}_A^*$  has the same first two moments as

$$\text{RMS}_A^* \sim \sigma_b^2 \chi_{\nu_1}^2 / \nu_1, \quad (26)$$

where

$$\nu_1 = \frac{[\sum_{ij} n_{ij}(1 - n_{ij}h_i)]^2}{\sum_{ij} n_{ij}^2(1 - n_{ij}h_i)}. \quad (27)$$

An unbiased estimator of  $\sigma_b^2$  is

$$\tilde{S}_b^2 = \left[ \frac{\sum_{ij} n_{ij}(1 - h_i)}{\sum_{ij} n_{ij}(1 - n_{ij}h_i)} \right] (\text{RMS}_A - \text{RMS}_B). \quad (28)$$

We will modify this estimator by truncating at zero:

$$S_b^2 \equiv \max(\tilde{S}_b^2, 0). \quad (29)$$

## 4 The Tolerance Limit Problem

With most of the distribution theory out of the way, we can now finally get to the statement of the problem to be addressed. Let  $w$  be an arbitrary known  $r \times 1$  vector, and define a random variable  $U$  such that

$$U \sim N(w^T \theta, \sigma_b^2 + \sigma_w^2). \quad (30)$$

We would like to construct a  $100\gamma\%$  lower confidence limit on the  $100(1 - \beta)$  percentile of  $U$  (where  $\beta = .9$  or  $\beta = .99$  for B- and A-basis values respectively). Upper and two-sided confidence limits can be defined similarly. Let  $\Phi(\cdot)$  denote the normal cdf, and define  $z_\beta$  so that

$$\Phi(z_\beta) = \beta. \quad (31)$$

We will determine a function  $K$  of  $\text{RMS}_A$  and  $\text{RMS}_B$  so that

$$\Pr(w^T \hat{\theta} - K \sqrt{\text{RMS}_A} \leq w^T \theta - z_\beta \sqrt{\sigma_b^2 + \sigma_w^2}) \approx \gamma. \quad (32)$$

We begin by evaluating

$$\text{Var}(w^T \hat{\theta}) = w^T (X^T X)^{-1} X^T \Sigma X (X^T X)^{-1} w, \quad (33)$$

using the identities

$$X^T X = \sum_i n_i x_i x_i^T \quad (34)$$

and

$$X^T \Sigma X = \sigma_w^2 \sum_i n_i x_i x_i^T + \sigma_b^2 \sum_{ij} n_{ij}^2 x_i x_i^T. \quad (35)$$

If  $S_b^2 = 0$ , then we will conclude that  $\sigma_b^2 = 0$ , and if  $S_b^2 = \infty$ , we will conclude that  $\sigma_w^2 = 0$ . In the latter case we are assured of being correct in our assumption, consequently the approximate tolerance limit which we will construct should be nearly exact in the limit of large between-batch variance. On the other hand, we can never conclude with certainty that  $\sigma_b^2 = 0$ , so the tolerance limit will provide only approximately the nominal confidence level  $\gamma$  when  $\sigma_b^2 = 0$ . If  $S_b^2 = 0$  and we assume that  $\sigma_b^2 = 0$ , then  $K \equiv K_0$ , where

$$K_0 = T_w^{-1} \left( \gamma, \frac{z_\beta}{\sqrt{w^T (X^T X)^{-1} w}} \right) \sqrt{w^T (X^T X)^{-1} w}. \quad (36)$$

If  $S_b = \infty$ , then  $\sigma_w^2 = 0$  and  $K \equiv \tilde{K}_1$ , where

$$\tilde{K}_1 = T_{n_1}^{-1} \left( \gamma, z_\beta \sqrt{c} \right) \sqrt{\frac{c \sum_{ij} n_{ij} (1 - h_i)}{\sum_{ij} n_{ij} (1 - n_{ij} h_i)}}, \quad (37)$$

and

$$c = w^T (X^T X)^{-1} \left[ \sum_{ij} n_{ij}^2 x_i x_i^T \right] (X^T X)^{-1} w. \quad (38)$$

We can write  $\tilde{K}_1$  as

$$\tilde{K}_1 = K_1 \sqrt{\frac{\sum_{ij} n_{ij} (1 - h_i)}{\sum_{ij} n_{ij} (1 - n_{ij} h_i)}}. \quad (39)$$

Note that  $K_0$  and  $K_1$  are of the form of normal distribution tolerance limit factors, with degrees of freedom for variance  $\nu_0$  and  $\nu_1$ , and effective number of observations for the mean  $[w^T(X^T X)^{-1}w]^{-1}$  and  $c^{-1}$ , respectively. For this reason we define

$$\eta_0^{-1} = w^T(X^T X)^{-1}w \quad (40)$$

and

$$\eta_1^{-1} = w^T(X^T X)^{-1} \left[ \sum_{ij} n_{ij}^2 x_i x_i^T \right] (X^T X)^{-1}w, \quad (41)$$

so that

$$K_0 = T_{\nu_0}^{-1}(\gamma, z_\beta \sqrt{\eta_0}) / \sqrt{\eta_0} \quad (42)$$

and

$$K_1 = T_{\nu_1}^{-1}(\gamma, z_\beta \sqrt{\eta_1}) / \sqrt{\eta_1}. \quad (43)$$

It is convenient to let

$$S_w^2 \equiv \text{RMS}_B, \quad (44)$$

since  $\text{RMS}_B$  is an unbiased estimator of  $\sigma_w^2$ . If we condition on  $\text{RMS}_A$  and  $\text{RMS}_B$ , solve for a 'tolerance limit factor'  $\tilde{K}$  which will depend on the unknown variances, replace  $z_\gamma$  and  $z_\beta$  with constants  $c_1$  and  $c_2$  to be determined, and replace the unknown variances with estimates, we end up with a tolerance limit factor

$$K \equiv c_1 \sqrt{\frac{S_{w^T \hat{\beta}}^2}{\text{RMS}_A}} + c_2 \sqrt{\frac{S^2}{\text{RMS}_A}}, \quad (45)$$

where

$$S^2 \equiv S_b^2 + S_w^2 \quad (46)$$

and

$$S_{w^T \hat{\beta}}^2 \equiv S_b^2 / \eta_1 + S_w^2 / \eta_0 \quad (47)$$

are estimators of  $\sigma_b^2 + \sigma_w^2$  and  $\text{Var}(w^T \hat{\theta})$ , respectively.

When  $S_b^2 = 0$ , we must have that

$$K_0 = c_1 / \sqrt{\eta_0} + c_2, \quad (48)$$



and when  $S_w^2 = 0$ ,

$$K_1 = c_1/\sqrt{\eta_1} + c_2. \quad (49)$$

These equation determine  $c_1$  and  $c_2$ , so that our tolerance limit factor is

$$K = \frac{\sqrt{\eta_0\eta_1}(K_1 - K_0)S_{w^T\hat{\beta}} + (K_0\sqrt{\eta_0} - K_1\sqrt{\eta_1})S}{\sqrt{\text{RMS}_A}(\sqrt{\eta_0} - \sqrt{\eta_1})}. \quad (50)$$

If  $S_b^2 = 0$ , then, for consistency,  $\text{RMS}_A$  should be set equal to  $\text{RMS}_B$  in the above formula. The tolerance limit factor is then

$$T = w^T\hat{\beta} - K\sqrt{\text{RMS}_A}, \quad (51)$$

where  $\sqrt{\text{RMS}_A}$  doesn't necessarily cancel with the denominator of  $K$ .

## 5 Examples

In order to illustrate the practicality of the above ideas, we discuss next two examples, both of which involve actual material strength data. The numerical results of this section were produced by a FORTRAN program which implements the methodology of this article, and which allows the specification of arbitrary nested mixed models. A preliminary version of this program is available from the author.

In the first example, 24 specimens from each of three batches of a graphite-epoxy composite material were tested in tension, with six specimens being broken at each of four temperatures. We would like to obtain a 95% two-sided confidence limit, as a function of temperature, for the 10th percentile of the population of a random strength measurement selected from a randomly chosen batch. We assume that the mean strength varies quadratically with temperature, even though a linear relationship fits nearly as well, in order to emphasize that we have the flexibility to choose any parametric model (provided, of course, that this model is linear in  $\theta$ ). The standard approach to this problem would begin with pooling the 18 observations at each temperature. Having ignored the batch effect, classical methodology can be used to calculate the desired confidence limit (e.g., Owen, 1968, pp. 462-463). The classical interval is displayed as the inner confidence interval in Figure 1. If we assume that the batches can be treated as if they were nested, then the methodology of this article leads to the

outer confidence interval in Figure 1. Note that this new interval is considerably wider, reflecting the uncertainty in the 10th percentile of strength due to between-batch processing variability in the composite. However, we have treated the batches as if they were nested, which is not the case for this experiment. This is an approximation which (on the basis of a formal hypothesis test) appears to be justified in this particular case.

In order to check whether, under the assumptions of the model, the confidence interval for the example in Figure 1 does indeed achieve nearly the nominal confidence level, 1000 random datasets  $Y = X\theta + \epsilon$  were generated for the appropriate  $X$  matrix and with normally distributed errors. The results of this simulation are not effected either by the particular value of  $\theta$  chosen or by  $\sigma^2 = \sigma_b^2 + \sigma_w^2$ ; however the intraclass correlation

$$\rho = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2} \quad (52)$$

is a nuisance parameter which can be expected to have some effect on the confidence level. Therefore, we chose to perform this simulation for  $\rho = 0, .25, .5, .75, 1$ , using the same pseudo-random numbers for each value of  $\rho$ . The extremes  $\rho = 0$  and  $\rho = 1$  correspond to no between-batch variability and no within-batch variability, respectively. The results of this simulation are displayed in Figure 2. Nominally, we would expect five percent of the replicates to fall in either tail; we can see from this figure that we nearly achieve this to within the error of the simulation, and that our results do not depend strongly on  $\rho$ .

As a second example, we consider data on the the pressures at which spherical Kevlar-epoxy pressure vessels failed by bursting (Gerstle and Kunz, 1983, p. 268). The data consist of 8 batches of sizes 6,5,5,2,3,2,5, and 1. We have no covariates in this example, so that we are actually dealing with an extremely unbalanced on-way random effects ANOVA model. In Figure 3, 90% two-sided confidence limits are given both for the case where the batches are pooled, and for the method of the present article. Necessarily, the interval which includes the between-batch variability is wider than the interval which ignores it. A simulation study was done for this example exactly as before, and the results, presented in Figure 4, are quite good. As simple as this example is, this analysis already goes beyond methodology in the

statistics literature, which is at present restricted to balanced models.

## 6 Conclusion

In this article, we have proposed an approach to determining approximate one-sided tolerance limits (or equivalently, approximate confidence limits on quantiles) for nested models with a single random effect and arbitrary fixed effects. This technique has potentially important applications to the characterization of composite material strength in the presence of between-batch processing variability. Two real-data examples illustrate the usefulness of the methodology, and the confidence levels in these examples have been shown to be close to the nominal levels, for all values of the unknown intraclass correlation, by simulation studies.

There are many possible directions for future work. Perhaps the most important of these is to work toward a better understanding of the approximation underlying the procedure. Experience with a variety of examples suggests that the approximation is a good one, but one can expect it to break down in certain situations, and we should try to find out what these situations are. Further generalizations of the work in this article are possible, at least formally. Possible generalizations include non-nested models, more than one random effect, two-sided tolerance limits, and non-normal random effects.

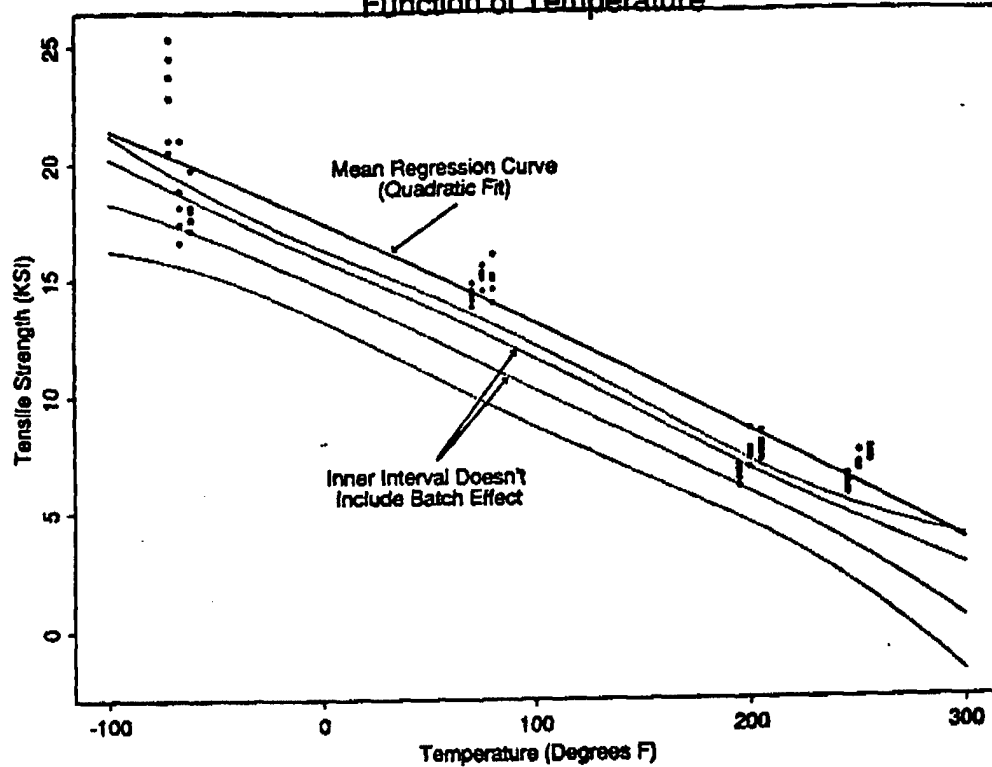
## References

- [1] Gere, J.M. and Timoshenko, S.P. (1984), *Mechanics of Materials*, Boston: Prindle, Weber & Schmidt.
- [2] Gerstle, F. P., Jr. and Kunz, S. C. (1983), "Prediction of Long-Term Failure in Kevlar 49 Composites", in *Long-Term Behavior of Composites*, T. K. O'Brien, ed., ASTM STP 813, American Society of Testing and Materials, Philadelphia, 263-292.
- [3] Mee, R. W. and Owen, D. B. (1983) "Improved Factors for One-Sided Tolerance Limits for Balanced One-Way ANOVA Random Model", *Journal of the American Statistical Association*, 78, 901-905.

- [4] Mil Handbook 5E (1987), *Metallic Components for Aircraft Structures*, Naval Publications and Forms Center, Philadelphia.
- [5] Mil Handbook 17C (1992), *Polymer Matrix Composites, Volume I: Guidelines*, Naval Publications and Forms Center, Philadelphia.
- [6] Owen, D. B. (1968), "A Survey of Properties and Applications of the Noncentral t-Distribution", *Technometrics*, 10, 445-478.
- [7] Seber, G. A. F. (1977), *Linear Regression Analysis*, John Wiley and Sons, New York.
- [8] Vangel, M. G. (1992), "New Methods for One-Sided Tolerance Limits for a One-Way Balanced Random-Effects ANOVA Model", *Technometrics* 34, 176.

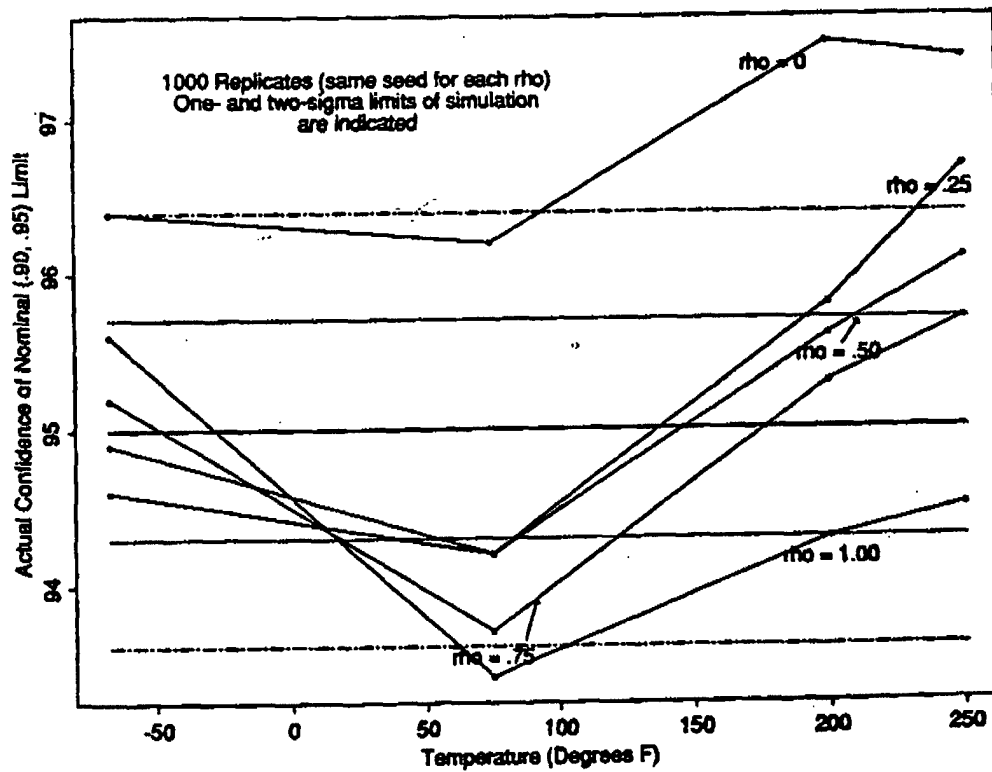
# Figure 1

90% Confidence Interval on 10th Percentile as a  
Function of Temperature

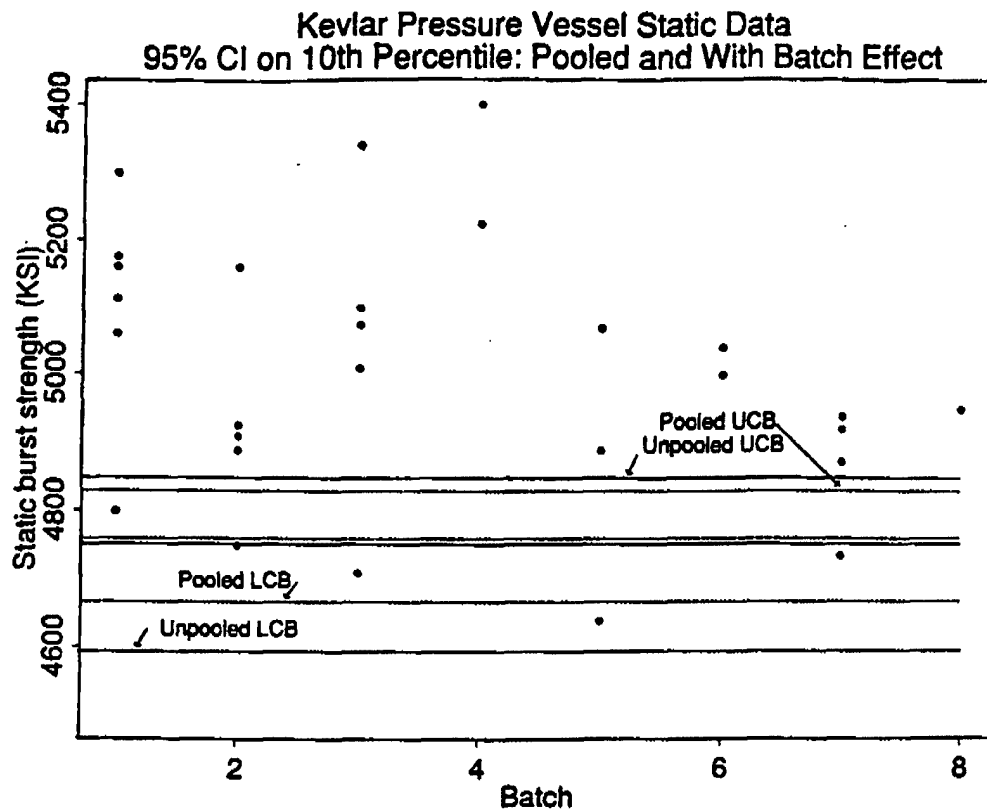


# Figure 2

Simulated Confidence for Strength/Temperature Regression

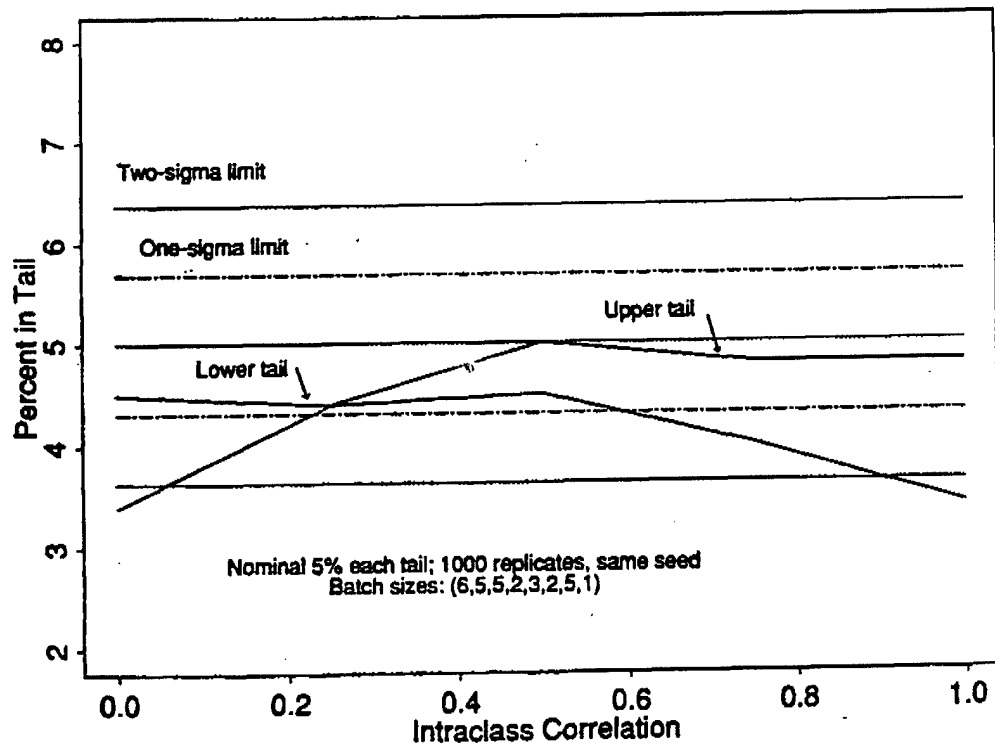


# Figure 3



# Figure 4

Simulated Coverage of 90% CI on 10th Percentile



DETERMINATION OF CAMOUFLAGE EFFECTIVENESS OF  
SMALL AREA CAMOUFLAGE COVERS (SACC)  
BY GROUND OBSERVERS USING THE METHOD OF LIMITS

GEORGE ANITOLE  
AND  
RONALD L. JOHNSON  
U.S. ARMY BELVOIR RESEARCH, DEVELOPMENT  
AND ENGINEERING CENTER  
FORT BELVOIR, VIRGINIA 22060-5606

CHRISTOPHER J. NEUBERT  
U.S. ARMY MATERIAL COMMAND  
ALEXANDRIA, VIRGINIA 22333-0001

ABSTRACT

The goal of this study was to evaluate the camouflage effectiveness of the Small Area Camouflage Cover (SACC) for a green and brown site. The SACC is designed to conceal individuals, small sized equipment and fighting positions. The test design consisted of the psychophysical Method of Limits to determine the just noticeable difference (JND) of each SACC. The JND is the distance that an observer has a fifty percent probability of reporting the seeing or not seeing of an object. Seven observers performed ten trials each, starting close to the SACC and walking back until they could not see the SACC, or starting at a distance where they could not see the SACC and walking toward it until the target was seen. The Student's T-Test was performed upon the JNDs to determine which SACCs were significantly ( $\alpha \leq 0.05$ ) more camouflage effective, individual differences for each observer, and differences in the two modes of target approach. This study presented a unique test design, and joined the expertise of an engineer, statistician, and psychologist.

1.0 SECTION I - INTRODUCTION

The SACC is a continuation of a program begun in 1986 by the Belvoir Research, Development and Engineering Center to develop an individual camouflage cover. The SACC is designed to provide protection from visual, near-infrared, and radar observation. It will conceal individual troops, or be attached together for use over weapon emplacements,

fighting positions, and supply caches. Each SACC weighs less than 518 grams (18 ozs.) and is small enough (2.13 X 1.37m) to be folded and fit into the pocket of a soldiers uniform.

The small size of the SACC precluded the type of usual range detection studies conducted in the past 1/ 2/ 3/. It was for that reason that the psychophysical Method of Limits 4/ was selected to determine the JND of each SACC. The JND is the distance that an observer has a fifty percent probability of reporting the seeing or not seeing of an object. This test design will be described in Section II.

## 2.0 SECTION II - EXPERIMENTAL DESIGN

### 2.1 Test Design

The Method of Limits is designed to determine the visual threshold JND of an object being viewed. This method is the only direct method of locating a threshold. The observers are started at either the far end of the observation path where the target SACC cannot be seen, or at the start of the observation path where the target SACC is easily seen. The observers know where the target is located in either situation. The observers proceed either toward the target or away from the target. They report when the target has just become visible or just disappeared from sight. This procedure was repeated ten times for each direction of target approach, and for each SACC at each of the two sites (See 2.3). The threshold has to be measured repeatedly, because its exact location varies from moment to moment. The marker nearest the threshold is recorded, and the mean distance determined. This mean is the threshold or JND for that observer for that particular SACC.

### 2.2 Test Targets

The test targets consisted of six candidate SACCs which have been coded to protect the identification of the individual manufacturers. The following is a brief description of the SACCs:

- o SACC A - Constructed of a polyester mesh material printed in the current woodland uniform pattern. It is not reversible, i.e. it does not have a different pattern or color on the other side.



o SACC B - Constructed of incised vinyl coated nylon scrim. It is reversible with a green pattern on one side and a brown pattern on the other side.

o SACC C - Constructed of variegated polyester knit. It is reversible with a green pattern on one side, and a brown pattern on the other side.

o SACC D - Constructed of incised vinyl coated spun bonded nylon material sewn to a black nylon scrim base cloth. The pattern is the same as found on the standard US woodland camouflage net. It is not reversible.

o SACC E - Constructed of a polyester mesh material which is not reversible. The pattern is made of large areas of black and green color.

o SACC F - Constructed of green and black dyed rip-stop nylon sewn in strips to a black mesh backing. It has 100 percent garnish cover and is not reversible.

## 2.3 Test Sites

There were a total of two sites used in this study. One site was inside the woods and consisted of a brown leafy background with tall trees. The other site was an open green field.

### 2.3.1 Brown Leafy

The brown leafy site was located at a bend in a straight road, and consisted of a small slope under a cover of large deciduous trees. The ground was covered with a thick mat of brown leaves. This site resemble what one would see inside the forest. The site offered a continuous line of sight of 880 feet. Forty two markers were placed, one every 20 feet starting 60 feet from the SACC, see Table 1.

### 2.3.2 Green Field

The green site was located on a hill at the Turner Drop Zone, Fort Devens, MA. The hill consisted of typical pasture grass and other growth, with a maximum height of about 2 feet. The site offered a continuous line of sight of 1560 feet. Thirty Five markers were placed, one every 40 feet starting 200 feet from the placement of the SACC, see Table 2.

TABLE 1

## DISTANCE OF MARKERS TO SACC FOR THE BROWN LEAFY SITE

<u>MARKER</u>	<u>DISTANCE IN FEET ALONG PATH FROM MARKER TO TARGET</u>	<u>MARKER</u>	<u>DISTANCE IN FEET ALONG PATH FROM MARKER TO TARGET</u>
1	60	22	480
2	80	23	500
3	100	24	520
4	120	25	540
5	140	26	560
6	160	27	580
7	180	28	600
8	200	29	620
9	220	30	640
10	240	31	660
11	260	32	680
12	280	33	700
13	300	34	720
14	320	35	740
15	340	36	760
16	360	37	780
17	380	38	800
18	400	39	820
19	420	40	840
20	440	41	860
21	460	42	880

TABLE 2  
DISTANCE OF MARKERS TO SACC FOR THE GREEN SITE

<u>MARKER</u>	<u>DISTANCE IN FEET ALONG PATH FROM MARKER TO TARGET</u>	<u>MARKER</u>	<u>DISTANCE IN FEET ALONG PATH FROM MARKER TO TARGET</u>
1	200	22	1040
2	240	23	1080
3	280	24	1120
4	320	25	1160
5	360	26	1200
6	400	27	1240
7	440	28	1280
8	480	29	1320
9	520	30	1360
10	560	31	1400
11	600	32	1440
12	640	33	1480
13	680	34	1520
14	720	35	1560
15	760		
16	800		
17	840		
18	880		
19	920		
20	960		
21	1000		

#### 2.4 Test Subjects

A total of seven military and civilians served as ground observers. All personnel had at least 20/30 corrected vision and normal color perception.

#### 3.0 SECTION III - RESULTS

Of the 6 SACCs studied, only 3 were effective in achieving a threshold JND. SACC B, which was reversible with a green and brown side, had a JND for the green and brown leafy sites. SACC C, also reversible, achieved a JND for the brown leafy site when the brown side was shown. Tables 3 and 4 summarize these results for the brown leafy site and the green site.

TABLE 3  
SUMMARY OF SACC DATA BROWN LEAFY SITE

SACC	JND	
	<u>IN</u>	<u>OUT</u>
A	No JND	No JND
B	JND	JND
C	JND	JND
D	No JND	No JND
E	No JND	No JND
F	No JND	No JND

No JND - SACC still visible at maximum range of 880 feet.  
Table 3 shows that only the SACC B and C had a JND.

TABLE 4  
SUMMARY OF SACC DATA GREEN SITE

SACC	JND	
	<u>IN</u>	<u>OUT</u>
A	No JND	No JND
B	JND	JND
C	No JND	No JND
D	No JND	No JND
E	No JND	No JND
F	No JND	No JND

No JND - SACC still visible at maximum range of 1,560 feet.  
Table 4 shows that only the SACC B had a JND.

The JNDs were calculated using two different mathematical approaches. The first calculated the mean of the ten trials approaching the target trials and the ten trials retreating from the target. This was done for all seven subjects and across all subjects. The results are found in Tables 5 through 10. The second type of mean threshold or JND was determined by adding each approaching and retreating JND and dividing by two. The overall mean across all observers was also determined. These results are seen in Tables 11 through 13. Table 14 contains the Student's T-Test 5/ for each of the three types of JNDs for the comparison of the SACC B vs. SACC C for the brown leafy site. Table 15 contains the probabilities for the Student's T-Test of Table 14.

TABLE 5

JUST NOTICEABLE DIFFERENCE, GROUND OBSERVERS APPROACHING  
TARGET, SACC B, BROWN LEAFY SITE

<u>OBSERVER</u>	<u>MEAN JND RANGE, FEET</u>	<u>STANDARD DEVIATION</u>
Anitole	152.00	12.29
Johnson	163.30	12.29
Bullock	220.60	19.47
Person	167.00	6.75
Wedemeyer	241.60	59.64
Sadley	224.70	11.18
Lyons	154.00	47.89
TOTAL:	189.03	46.33

TABLE 6

JUST NOTICEABLE DIFFERENCE, GROUND OBSERVERS APPROACHING  
TARGET, SACC C, BROWN LEAFY SITE

<u>OBSERVER</u>	<u>MEAN JND RANGE, FEET</u>	<u>STANDARD DEVIATION</u>
Anitole	263.00	25.41
Johnson	175.70	28.93
Bullock	186.90	42.23
Person	185.00	48.59
Wedemeyer	275.30	76.89
Sadley	226.10	47.74
Lyons	126.00	15.78
TOTAL:	205.43	65.09

TABLE 7

JUST NOTICEABLE DIFFERENCE, GROUND OBSERVERS APPROACHING  
TARGET, SACC B, GREEN SITE

<u>OBSERVER</u>	<u>MEAN JND RANGE, FEET</u>	<u>STANDARD DEVIATION</u>
Anitole	440.00	35.28
Johnson	371.90	61.09
Bullock	401.60	150.38
Person	298.00	56.13
Wedemeyer	773.00	176.03
Sadley	731.60	222.62
Lyons	459.40	260.82
TOTAL:	496.50	228.93

TABLE 8

JUST NOTICEABLE DIFFERENCE, GROUND OBSERVERS RETREATING FROM  
TARGET, SACC B, BROWN LEAFY SITE

<u>OBSERVER</u>	<u>MEAN JND RANGE, FEET</u>	<u>STANDARD DEVIATION</u>
Anitole	123.00	13.98
Johnson	180.70	16.64
Bullock	209.30	21.34
Person	200.00	9.43
Wedemeyer	219.40	38.87
Sadley	224.00	18.30
Lyons	182.00	35.52
TOTAL:	191.20	39.63

TABLE 9

JUST NOTICEABLE DIFFERENCE, GROUND OBSERVERS RETREATING FROM  
TARGET, SACC C, BROWN LEAFY SITE

<u>OBSERVER</u>	<u>MEAN JND RANGE, FEET</u>	<u>STANDARD DEVIATION</u>
Anitole	224.00	25.91
Johnson	190.80	33.73
Bullock	176.00	40.79
Person	230.00	46.90
Wedemeyer	252.30	64.34
Sadley	221.30	61.98
Lyons	172.00	30.48
TOTAL:	209.49	51.93

TABLE 10

JUST NOTICEABLE DIFFERENCE, GROUND OBSERVERS RETREATING FROM  
TARGET, SACC B, GREEN SITE

<u>OBSERVER</u>	<u>MEAN JND RANGE, FEET</u>	<u>STANDARD DEVIATION</u>
Anitole	374.00	44.27
Johnson	309.90	47.69
Bullock	392.30	174.39
Person	278.00	51.16
Wedemeyer	658.00	166.65
Sadley	660.50	241.86
Lyons	538.60	143.88
TOTAL:	458.76	202.53

TABLE 11

APPROACHING PLUS RETREATING JND DIVIDED BY TWO,  
GROUND OBSERVERS, SACC B, BROWN LEAFY SITE

<u>OBSERVER</u>	<u>MEAN JND RANGE, FEET</u>	<u>STANDARD DEVIATION</u>
Anitole	137.50	10.41
Johnson	172.00	9.00
Bullock	214.95	20.25
Person	183.50	7.09
Wedemeyer	238.50	49.19
Sadley	224.35	14.28
Lyons	168.00	40.79
TOTAL:	191.25	40.81

TABLE 12

APPROACHING PLUS RETREATING JND DIVIDED BY TWO,  
GROUND OBSERVERS, SACC C, BROWN LEAFY SITE

<u>OBSERVER</u>	<u>MEAN JND RANGE, FEET</u>	<u>STANDARD DEVIATION</u>
Anitole	243.50	24.16
Johnson	183.25	30.38
Bullock	181.45	41.49
Person	207.50	47.51
Wedemeyer	263.00	70.53
Sadley	223.70	54.79
Lyons	149.00	21.83
TOTAL:	207.46	56.42

TABLE 13

APPROACHING PLUS RETREATING JND DIVIDED BY TWO,  
GROUND OBSERVERS, SACC B, GREEN SITE

<u>OBSERVER</u>	<u>MEAN JND RANGE, FEET</u>	<u>STANDARD DEVIATION</u>
Anitole	407.00	33.02
Johnson	340.90	52.05
Bullock	396.95	156.73
Person	288.00	48.03
Wedemeyer	715.50	171.32
Sadley	696.05	231.48
Lyons	499.00	196.43
TOTAL:	477.62	211.23

TABLE 14

STUDENTS T FOR THE COMPARISON OF THE MEAN JND OF  
SACC B vs. SACC C IN BROWN LEAFY SITE  
FOR EACH GROUND OBSERVER

OBSERVER	APPROACHING JND	RETREATING JND	APPROACHING PLUS RETREATING JND DIVIDED BY TWO
Anitole	-12.44	-10.85	-12.74
Johnson	- 1.25	- 0.85	- 1.12
Bullock	2.29	2.29	2.29
Person	- 1.16	- 1.98	- 1.58
Wedemeyer	- 1.10	- 1.38	- 1.22
Sadley	- 0.09	0.13	0.04
Lyons	1.76	0.68	1.38
TOTAL:	- 1.72	- 2.34	- 2.08

A negative sign means that the mean JND for SACC B is smaller than the mean JND for SACC C.

TABLE 15

PROBABILITIES OF THE STATISTICAL COMPARISON OF  
THE MEAN JND OF SACC B vs. SACC C,  
BROWN LEAFY SITE FOR EACH GROUND OBSERVER

OBSERVER	APPROACHING JND	RETREATING JND	APPROACHING PLUS RETREATING JND DIVIDED BY TWO
Anitole	*0.000	*0.000	*0.000
Johnson	0.228	0.407	0.276
Bullock	**0.039	**0.034	**0.034
Person	0.261	0.063	0.132
Wedemeyer	0.288	0.183	0.237
Sadley	0.929	0.896	0.971
Lyons	0.096	0.508	0.210
TOTAL:	0.088	**0.021	**0.039

\* Significant at  $\alpha \leq 0.001$

\*\* Significant at  $\alpha \leq 0.05$

#### 4.0 SECTION IV - DISCUSSION

Tables 3 and 4 show that the SACCs B and C achieved a JND threshold for the brown leafy site. The reversible



green side of the SACC B was the only one to achieve a JND at the green site. No JND at the brown leafy site means that the SACC was seen at the maximum range of 880 feet. The maximum range at the green site was 1,560 feet. An inspection of Tables 5-13 shows that there are large individual between observers, but each observer had a fairly constant approaching and retreating threshold. That is the approaching and retreating threshold are very close to each other. Thus the use of the Method of Limits was very successful in determining the threshold of the SACCs in terms of range. Table 14 indicated that SACC B had a smaller JND 13 out of 21 times. This difference was significant for the retreating JND and the averaged approach retreating JNDs,  $\alpha \leq 0.05$  (Table 15).

## 5.0 SECTION V - SUMMARY AND CONCLUSIONS

A total of six SACCs were evaluated by ground observers to determine the range of effectiveness JND. The JND is defined as the distance at which a target starts to appear or blend into the background. The ground observers started from both the beginning and end of an observation path with surveyed stations. The observation station nearest the JND was recorded. The test targets consisted of six candidate SACCs which were coded to protect the identification of the individual manufacturer. The following conclusions were determined:

- a. SACC B had JNDs for both the brown leafy site and the green site.
- b. SACC C had a JND for the brown site.
- c. The JND threshold varied greatly from observer to observer, but was constant within each observer for the approaching and retreating threshold determination.
- d. The Method of Limits is a good test design to obtain range data for small sized test items.

## REFERENCES

1. Anitole, George, Johnson, Ronald L., and Neubert, Christopher, Evaluation of Camouflage Paint Gloss vs. Detection Range, Proceedings of the Thirty-Third Conference on the Design of Experiments In Army Research, Development and Testing, 1987.

2. Anitole, George, Johnson, Ronald L., and Neubert, Christopher, Determination of Monotone and Camouflage Patterned Five-Soldier Crew Tests by Ground Observers, Proceedings of the Thirty-Fourth Conference on the Design of Experiments In Army Research, Development and Testing, 1988.
3. Anitole, George, Johnson, Ronald L., and Neubert, Christopher, Determination of Detection Range of CUCV Trucks with Monotone and Camouflage Pattern Tarpaulins, Proceedings of Test Technology Symposium II, April 1989.
4. Woodworth, Robert S. and Schlosberg, Harold, Experimental Psychology, Henry Holt and Company, Inc., August 1956.
5. Natrella, Mary G., Experimental Statistics, National Bureau of Standards Handbook 91, U.S. Department of Commerce, Washington, D.C., 1966.

# AN EXPLORATORY ALGORITHM FOR THE ESTIMATION OF MODE LOCATION AND NUMEROSITY IN MULTIDIMENSIONAL DATA\*

Marc N. Elliott and James R. Thompson  
Rice University, Houston, Texas 77251-1892

## Abstract

Many advocate using nonparametric density estimation as a tool for mode estimation. While this approach may be appropriate in the univariate and bivariate cases, it can be quite inefficient in higher dimensional situations. A nonparametric algorithm is presented which determines the number and location of modes in a multidimensional data set. The procedure can be used in data exploration and can also automatically and nonparametrically test for multimodality. Several applications are discussed. In particular, it is demonstrated that the Fisher-Anderson iris data, which contains 3 species, has 4 modes.

Consider the problem of the exploratory analysis of a data set with four or more dimensions, relatively few observations, and large differences in scale. Since there is much "empty space" in high dimensional data sets, a good first step would be to find *modes*, local maxima of probability density. These modes could then serve as "base camps" for the further exploration of the structure of the data.

It would not be easy to enumerate and locate these modes using standard nonparametric density estimation techniques. Large differences in scale are a real difficulty for kernel density estimation, given that choice of kernel size is a critical problem. Furthermore, high dimensionality is quite problematic for standard methods of nonparametric density estimation. Such techniques often require unreasonably many observations for high dimensional problems.

Imagine now a data set that one would never even consider analyzing with standard techniques: 200 12-dimensional observations. According to Silverman(1986), one would need 842,000 observations just to get good density estimates near the mode of a standard multivariate normal distribution in 10 dimensions.

Most researchers would probably discard all but two or three dimensions or would only collect two or three variables in the first place. But throwing out variables is throwing away information that can help locate modes. Mode estimation should become easier in high dimensions, if one can harness the additional information without being swamped by the empty space. We have developed

\*This research was supported, in part, by the Army Research Office (Durham) under DAAL-03-91-G-0210 at Rice University.

an algorithm that can do so, automatically.

## METHOD

Before analysis, data is standardized dimension by dimension, so that each dimension has mean zero and variance one. After analysis, the "z-transform" is reversed, returning the units to those in the raw data. The standardization is done to prevent scale differences among dimensions from unduly influencing the analysis. If one has *a priori* knowledge that the dimensions are naturally of comparable scale, one might want to omit this standardization.

The algorithm itself consists of three stages. In the first stage, data is condensed locally. In the second stage, the condensed representation is used to construct "best" models for different numbers of modes. In the third stage, one of those models is selected through a hierarchical test.

### Condensing Data Locally

The central element of this stage is the *Mean Update Algorithm* (MUA), explicated in previously published work (Boswell, 1983; Thompson & Tapia, 1990). The MUA is an algorithm for moving toward local centers of probability density from any starting point in multi-dimensional space. It has the property that it is not thwarted by high dimensionality. An outline of the algorithm follows:

```
    Select a starting point, x
    Select a # of nearest neighbors parameter, k
        update=True
        While update=True
            Select the k nearest neighbors of x in terms of Euclidian distance
            x'= the dimension-by-dimension mean of these k neighbors
            If x'=x then update=False
        End
    Output x
```

It should be noted that Boswell established that the MUA always terminates.

The MUA is used in the following manner in order to locally condense a multidimensional data set of size  $n$ :

```
    Select a parameter, mm << n
        m = n
        While m > mm
            Run the MUA n times in parallel, using each data point as a starting point
            Record the n outputs as the new data set
            m = the number of different points in the data set
        End
    Output the condensed data set of size n
```

The parameter  $mm$  represents the largest number of modes that one wants to consider as a possibility. Excessively large values of  $mm$  result in inefficiency. Often  $mm=5$  is a good default.

While the MUA itself always terminates, the process of condensing the data set using the MUA can "stall". On rare occasions, the number of different points in the data set,  $m$ , will not decrease from one iteration of the condensation algorithm to the next. In this case, the number of different points in the data set is reduced to  $mm$  in a manner similar to that used for reducing the number of modes in the following stage.

The selection of the number of nearest neighbors parameter,  $k$ , is very important. A value that is too large tends to combine distinct modes. A value that is too small is inefficient and also has a slight tendency to fracture unitary modes. If one is running the algorithm with a human observer, it is easy to adjust  $k$  to an appropriate level. The practical advantages of an automated algorithm, however, demand that we find a workable automatic procedure for selecting  $k$ .

Mack and Rosenblatt (1979) showed that  $k=cn^{4/(4+d)}$  has optimal properties for nearest neighbor density estimation in dimension  $d$ . Wong and Lane (1983) recommended  $k=n^{1/2}$  for clustering multidimensional data with nearest neighbor techniques. We found that  $k=n^{1/2}$  works well in our application. Note that our value approximates that of Mack and Rosenblatt for  $d=3$  to 8. If one knows that one's data is *extremely* kurtotic, one might want to consider reducing  $k$ , since highly kurtotic multidimensional data exhibit very little clustering. Nevertheless, the given value of  $k$  is very broadly applicable, even with kurtotic data, as will be seen.

### Constructing Models

Consider the output of the first stage. It is a data set containing  $n$  points,  $mm$  of which are different. Alternatively, it consists of  $mm$  values, each of which has  $n_i$  replications, such that

$$i=1, \dots, mm \text{ and } \sum_{i=1}^{mm} n_i = n.$$

One can consider these  $mm$  values to represent point estimates of potential modes, since they are the locations in multidimensional space to which the MUA was drawn. Furthermore, the  $n_i$  reflect the relative density of the space surrounding each potential mode, since they ultimately reflect the number of points in the original data set that were drawn to the potential mode. One might then consider this representation to be a  $mm$  mode model of the data set, complete

with information relating to the local density surrounding each potential mode.

The  $mm$  mode model is then used to construct models consisting of 1, 2, . . .  $mm$  modes in the following manner:

```

For i= mm to 1
  Find 2 nearest remaining modes, in terms of Euclidian distance
  Replace the 2 modes with 1; a dim. by dim. average, weighted by  $n_i$ 's
  Let the  $n_i$  of the new mode be the sum of those combined to form it
End

```

### Hierarchical Test of Modes

At this point there exist  $mm$  models of the data, consisting of 1, 2, . . .  $mm$  modes. A hierarchical test procedure is then used to chose among the models:

```

i=0
rejectnull=TRUE
While rejectnull=TRUE
  i=i+1
  Test  $H_0$ : i modes vs.  $H_1$ : i+1 modes
  If BINOMIAL TEST OF BIMODALITY (BTOB) fails to reject, rejectnull=FALSE
End
Conclude i modes

```

The procedure is based on a test we call the *Binomial Test of Bimodality*. This test was developed to determine whether a given region of multidimensional space is better represented by one or two modes. If one views the process of combining modes as outlined above in reverse order, it can be seen as successive splitting of one mode in the set into two new modes. The BTOB can be applied to the region of space involving the mode proposed to split. Incorporating this into the hierarchical procedure outlined above allows one to decide upon the optimal number of modes and hence the best representation.

The principle of the BTOB is as follows. Let us call the mode that the algorithm may split  $M1$ . Let us call the two modes it splits into  $M2a$  and  $M2b$ . Spatially,  $M1$  will lie between  $M2a$  and  $M2b$ . If the alternative hypothesis of two modes is true, the density near  $M2a$  and  $M2b$  should be high relative to that near  $M1$ . If the null hypothesis of one mode is true, the density near  $M1$  should be high relative to that near  $M2a$  and  $M2b$ . The BTOB will simply be a one-tailed test of the alternative hypothesis that the density near  $M2a$  and  $M2b$  is higher than that near  $M1$  versus the conservative null hypothesis that the density near  $M1$  is equal to that near  $M2a$  and  $M2b$ .

The BTOB proceeds as follows:  $D$ -dimensional spheres are constructed around each of the points  $M1$ ,  $M2a$ , and  $M2b$ . The ratios of the radii (and hence the ratios of the volumes) are fixed such that the radii are proportional to  $n_i$ , the measure of relative local density near each of the three points. Given the process of mode splitting, this ensures that the volume of the sphere surrounding  $M1$  is equal to the sum of the volumes of the spheres surrounding  $M2a$  and  $M2b$ . Thus, under the null hypothesis, an equal number of points from the original data set would be expected in the first sphere as in the second two combined. Under the alternative, more would be expected in the second two. Initially the radii are scaled such that they just intersect. They are then slowly enlarged proportionately until the number of points captured in the three spheres reaches 25% of the  $n_i$  for  $M1$ . Under the null hypothesis, the number of points captured in the  $M1$  sphere is distributed binomially, with the number of trials being the total number of captured points and the probability of success being .5. An exact binomial test of this null hypothesis is then performed, using a nominal type I error rate of .05

## RESULTS

### Simulations

The algorithm was tested with multimodal simulated data generated from mixture densities. The algorithm was tested with 25 simulations under each of 72 sets of conditions. The sets of 72 conditions consisted of all possible combinations of four factors.

#### Factors Examined

The first factor examined was the *type of distribution* used for the unimodal densities comprising the mixture. Uncorrelated multivariate normal, slightly correlated multivariate normal, and uncorrelated multivariate  $t$  with 3 degrees of freedom were used. The  $t$  distribution was used to illustrate the performance of the algorithm under conditions of high kurtosis. The slightly correlated distribution had a correlation matrix with values ranging from .1 to .2.

The second factor examined was the *dimensionality* of the data set. Data sets of 4, 8, and 12 dimensions were used. It seems unnecessary to investigate the algorithm's performance in lower dimensionalities, as adequate techniques for mode estimation already exist for such cases.

The third factor examined was the *number of modes* in the mixture density. Densities with 1, 2, 3, and 4 modes were used. For all simulations, the maximum number of modes parameter, *mm*, was set to 5, so that the algorithm only considered the possibility of 1 to 5 modes.

The fourth factor examined was *sample size*. Very modest sample sizes of 100 and 200 were used. Larger sample sizes were used in an application that follows.

Figures 1a-1d illustrate the analysis of 200 observations from a 12-dimensional data set with 4 modes. The figures show 2 of the 12 dimensions of the data set.

### **Construction of the Densities**

The centers of the constituent densities of the mixture densities were determined by sampling from an uncorrelated multivariate normal distribution. The proportion of points allocated to each constituent density was as follows: .75 and .25 for 2 mode densities; .45, .35, and .20 for 3 mode densities; and .40, .24, .19, and .17 for 4 mode densities.

### **Distance Between Modes**

The amount of distance between the modes in the multimodal mixture density is very important in determining how difficult the mode estimation problem is. The metric used was the *Euclidian distance between the two nearest modes of the multimodal mixture, divided by the number of dimensions*. The minimum distance between modes was found to reflect the difficulty of the problem more accurately than the average or median distance between modes. These distances were divided by the number of dimensions so that they corresponded to the amount of information present per dimension. The units in which the distances are measured are *standard deviations of the unimodal densities comprising the mixture*.

Two aspects of the algorithm's performance were measured. The first was the accuracy of the algorithm in determining the correct number of modes. The second was the accuracy of the estimates of mode location.

### **Determining the Correct Number of Modes**

The criterion used to assess performance in this regard was the smallest amount of distance between modes (as defined above) that resulted in the algorithm correctly determining the number of modes in 20 or more of the 25 simulations. This quantity will be called the *Separation Needed for Accurate Mode Counts (SNAMC)*.



Table 1 reports SNAMC for Uncorrelated MV Normal Mixtures.

**Table 1: Uncorrelated MV Normal Mixtures (SNAMC)**

	N=100			N=200		
Dim.	2 modes	3 modes	4 modes	2 modes	3 modes	4 modes
4-D	3.75	5.3	4.3	3.0	4.0	3.2
8-D	2.75	3.7	3.6	2.25	2.6	2.2
12-D	2.0	3.7	3.4	2.0	2.4	1.8

Several trends are apparent in Table 1. First, performance improves with dimensionality. As dimensionality increases, less information is needed per dimension in order to achieve a given level of performance. Second, performance is much higher with a sample size of 200 than with a sample size of 100. This is promising for larger samples. No trends are apparent with respect to numerosity of modes. Finally, it should be noted that these levels of performance are quite good. Figure 1a shows 2 dimensions of a 200 observation, 12-dimensional mixture of 4 uncorrelated multivariate normal distributions with a minimum separation between modes of 2 standard deviations per dimension, corresponding to the SNAMC above. As can be seen, the overlap in distributions is substantial.

The same amount of separation between modes (standardized by standard deviation) that constituted the SNAMC with uncorrelated multivariate normal mixtures was used in simulations involving mixtures of uncorrelated multivariate t distributions with 3 degrees of freedom. In all cases, the correct number of modes were declared in 20 or more of the 25 simulations. Thus it appears that fairly substantial kurtosis does not adversely affect the performance of the algorithm.

The effects of correlation in the data are apparent in Table 2.

**Table 2: SNAMC of Correlated and Uncorrelated MVN mixtures (N=100)**

	Uncorrelated			Slightly Correlated		
Dim.	2 modes	3 modes	4 modes	2 modes	3 modes	4 modes
4-D	3.75	5.3	4.3	4.5	6.0	4.8
8-D	2.75	3.7	3.6	3.5	4.2	4.0
12-D	2.0	3.7	3.4	3.0	4.2	4.0

It is clear that more separation between modes is required for correlated data. This is unsurprising, since correlation among the variables means that less total information is present.

Unimodal distributions do not involve separation of modes, so the criterion used for unimodal densities was the proportion of the cases in which their unimodality was identified. This occurred in 425 of 450 runs (94.4%), which corresponds closely to the nominal type I error rate of 5%. In each of the cases in which multimodality was falsely claimed, the algorithm declared 2 modes.

### Accuracy of Estimates of Mode Location

Location accuracy was examined for data sets with separation corresponding to the SNAMC listed above. It was measured, mode-by-mode, only for those cases in which the number of estimated modes was correct. The criterion used to assess performance in this regard was based on a MSE criterion. The MSE of the estimate of a given mode was defined as follows:

Let the true mode location be  $(x_1, \dots, x_p)$  in  $p$  dimensions

Let the estimated location be  $(x^*_1, \dots, x^*_p)$

$$MSE = \frac{1}{p} \sum_{i=1}^p (x_i - x^*_i)^2$$

The average MSE (AMSE) for each of the modes was the criterion used. Note that the MSE will vary with the density surrounding each mode (or the mixture proportions in the case of mixture densities). Also note that the best expected AMSE that could be achieved for a given mode in a mixture density would be the AMSE of the mean of only those points from the correct subdistribution. This AMSE would be  $1/(prop*n)$  for subdistributions with unit variance and  $prop*n$  points. The units of the AMSE will be the variance of the subdistributions. AMSE's for sample sizes of 100 are in Table 3. As can be seen in the table, distributional form and dimensionality do not affect the accuracy of estimation of mode location *once the modes are separated sufficiently for their number to be accurately assessed*. How much separation is required for this to occur is affected by dimensionality and correlation, as was seen before. The AMSE's are roughly inversely proportional to the number of points in the subdensity. Similarly, the MSE's with a sample size of 200 are approximately half of what they are with a sample size of 100.

Finally, it should be noted that the accuracy of the location estimates is quite high. The AMSE's reported for a sample size of

200 imply that the standard deviations of estimates of mode location are  $1/8$  to  $1/4$  the standard deviations of the subdistributions, depending on the size of the mode.

**Table 3: AMSE of Mode Location Estimates, N=100**  
**2 Modes                      3 Modes                      4 Modes**

Dist'n	Dim.	.75	.25	.45	.35	.20	.40	.24	.19	.17
Normal Uncorr	4-D	.045	.088	.046	.070	.123	.048	.092	.121	.144
	8-D	.050	.100	.051	.063	.102	.056	.099	.106	.112
	12-D	.067	.101	.053	.059	.089	.065	.097	.098	.104
Normal Corr	4-D	.037	.078	.049	.065	.116	.067	.079	.143	.150
	8-D	.049	.100	.046	.059	.112	.055	.098	.090	.123
	12-D	.056	.097	.051	.052	.096	.053	.092	.094	.108
T(3) Uncorr	4-D	.030	.076	.060	.060	.084	.063	.100	.153	.107
	8-D	.046	.086	.060	.076	.100	.083	.093	.120	.121
	12-D	.058	.088	.080	.077	.105	.088	.105	.110	.137

### Applications

It is difficult to find real-world data sets which have been collected with a large number of variables, simply because researchers do not tend to collect data for which they lack effective analytical tools. We were able to find two 4-dimensional data sets to which we could apply the algorithm. The first, a set of observations regarding shell penetration behind an armor plate, was kindly supplied by Dr. Malcolm S. Taylor of the Army Research Laboratory in Aberdeen, Maryland. The second is the well-known Fisher-Anderson iris data.

#### Shell Penetration Behind an Armor Plate

The data is a collection of 944 observations regarding shell penetration behind an armor plate. The identity of the variables is classified, but it may be noted that the means of the variable are of the orders  $10^2$ ,  $10^1$ ,  $10^1$ , and  $10^{-1}$ . Their standard deviations are of the orders  $10^2$ ,  $10^1$ ,  $10^1$ , and  $10^0$ , correspondingly. Two of the four dimensions are pictured in Figure 2a. In less than 2 minutes, the algorithm was able to find 2 modes, as pictured in Figure 2b. The presence of 2 modes was confirmed by sources familiar with the data set.

There are additional uses that may be made of this data set. One could transfer the data from one 944 X 4 matrix to another, reading down the columns of the first matrix, but entering across the rows of the second matrix. This would result in a new data set of 944 observations in four dimensions. Three modes, corresponding to the large-scale variable, the two medium-scale variables, and the small-scale variable of the original data would be created by the natural differences in scale of the original variables. These modes would contain 25%, 50%, and 25% of the data, respectively. This new data set would thus be a stringent test of an automatic algorithm's ability to cope with vast differences in scale. Figure 3a shows a representation of the altered data set. Figure 3b shows the 3 modes which the algorithm successfully located. Finally, Figures 3c and 3d enlarge the scale to reveal the "micro-universes" in which two of the modes reside.

#### Fisher-Anderson Iris Data

The Fisher-Anderson iris data is a well known set consisting of 150 observations of 4 variables. The 150 observations consist of 50 observations of each of 3 Iris species: *I. setosa*, *I. versicolor*, and *I. virginica*. The variables measured were sepal length, sepal width, petal length, and petal width. Table 4 lists the means of the species on each of the variables.

**Table 4: Mean Characteristics of Irises, by Species**

Species	Sepal Length	Sepal Width	Petal Length	Petal Width
<i>Setosa</i>	5.006	3.428	1.462	.246
<i>Versicolor</i>	5.936	2.770	4.260	1.326
<i>Virginica</i>	6.588	2.974	5.552	2.026

*I. setosa* is distinguished from the other two species relatively easily. *I. versicolor* and *I. virginica*, however, are very difficult to distinguish. In fact, the proximity of the means of the two distributions relative to their variances, in combination with the very high degree of correlation present in the variables, has made the data set famous for its difficulty.

Intuitively, one would assume that the density consisting of a mixture of three species would have three modes. Indeed, it has long been the aim of mode-finders to demonstrate that their algorithm declares the Fisher-Anderson data to contain 3 modes, as opposed to the 2 modes (*setosa*, *versicolor/virginica*) that are usually found. When we ran the algorithm on the data set, it declared 4 modes, as is shown in Table 5.

**Table 5: Estimated Modes of Iris Data**

Mode Name	Sepal Length	Sepal Width	Petal Length	Petal Width
<i>Setosa</i>	4.992	3.411	1.462	.225
<i>Versicolor*</i>	5.642	2.696	4.101	1.267
<i>Versi/Virgin</i>	6.249	2.893	4.837	1.591
<i>Virginica*</i>	6.762	3.067	5.589	2.218

The algorithm found one mode at the mean of the *setosa* distribution. This location was estimated with extraordinary accuracy. It then found 3 modes for the species *versicolor* and *virginica*: : one right between the modes of the two species, and two to the outsides of the species means. All three estimates fell on a line.

We decided to investigate whether three mode actually exist for the two species. We did this by projecting the *versicolor* and *virginica* observations onto the line determined by the three estimates of mode location. The gaussian-smoothed histogram of these projections (Figure 4a) strongly supports this conjecture of three modes from two species. Similar results were obtained by projecting onto the line determined by the *versicolor* and *virginica* means (Figure 4b). It seems likely that this local trimodality is caused by overlapping distributions, and that the Fisher-Anderson iris data, while consisting of 3 species, contains 4 modes.

Fig. 1a: Simulated Data, 200 12-d Obs.

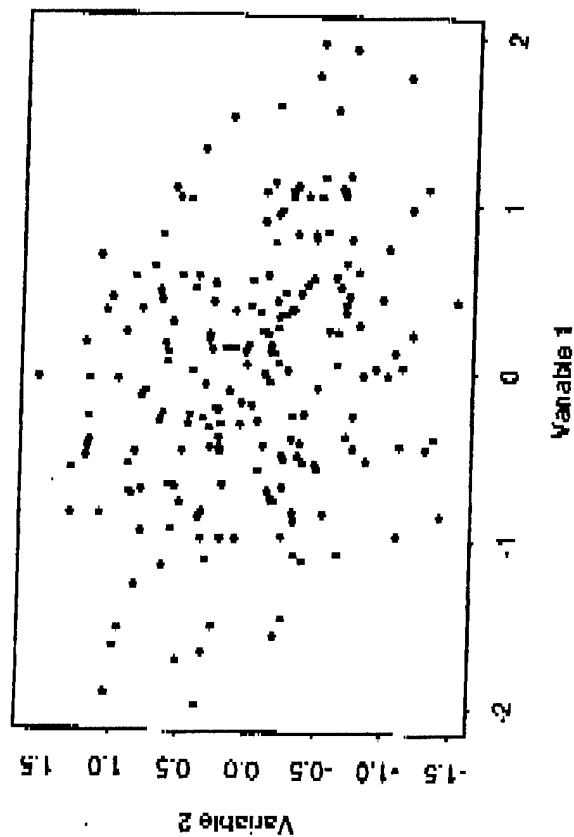


Fig. 1b: Data After 1 Condensation

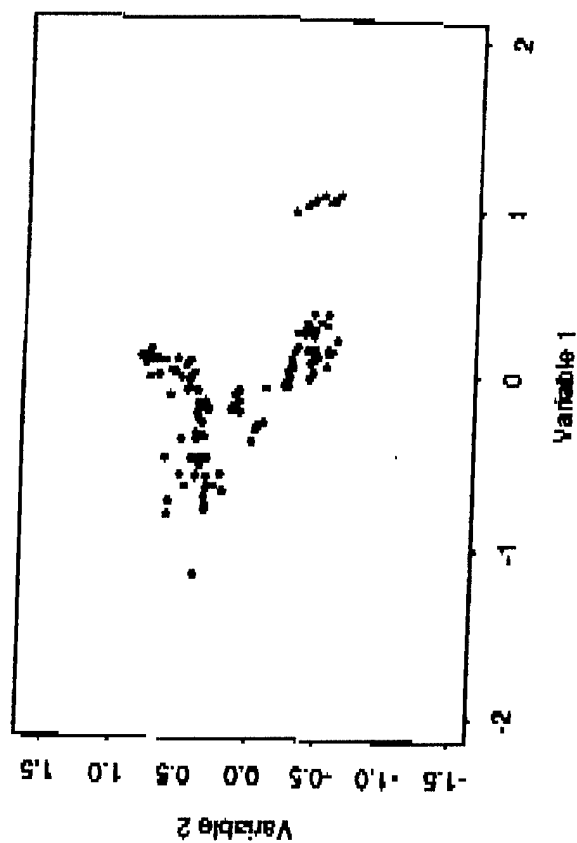


Fig. 1c: Data After 2 Condensations

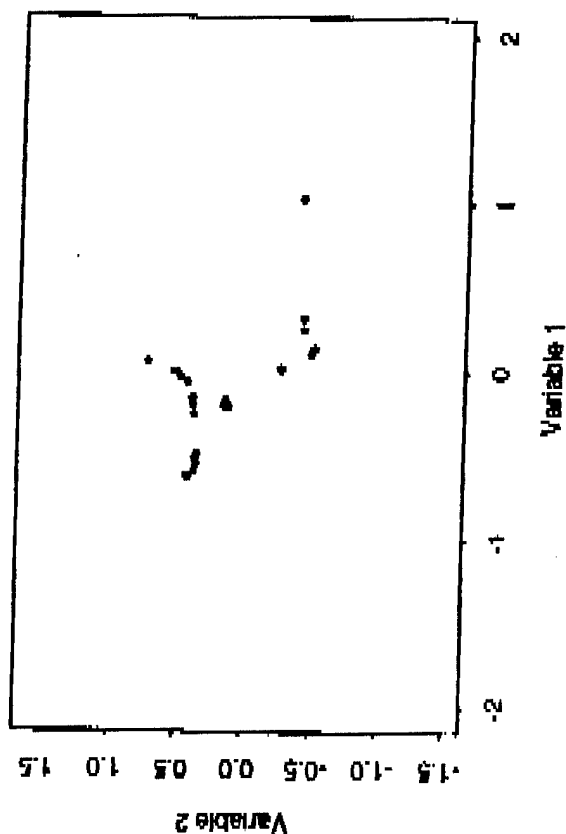


Fig. 1d: Final Estimates of 4 Modes

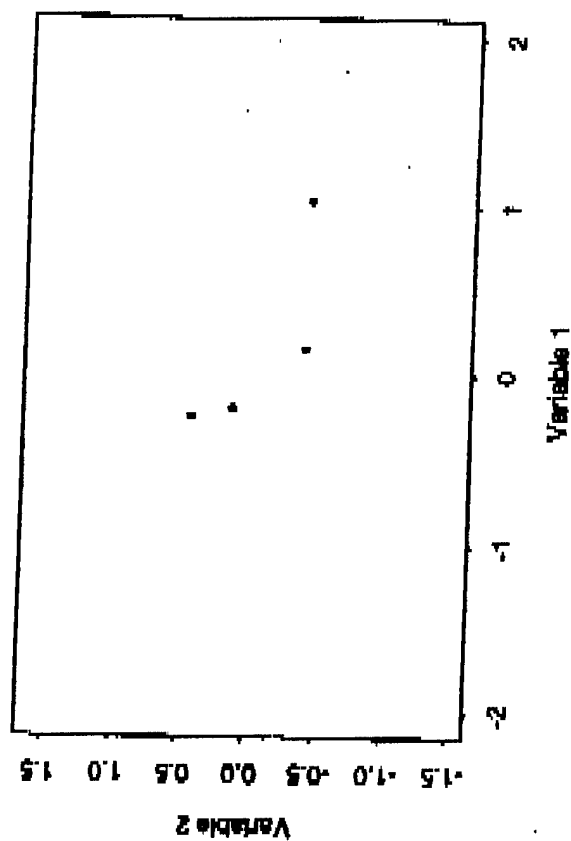
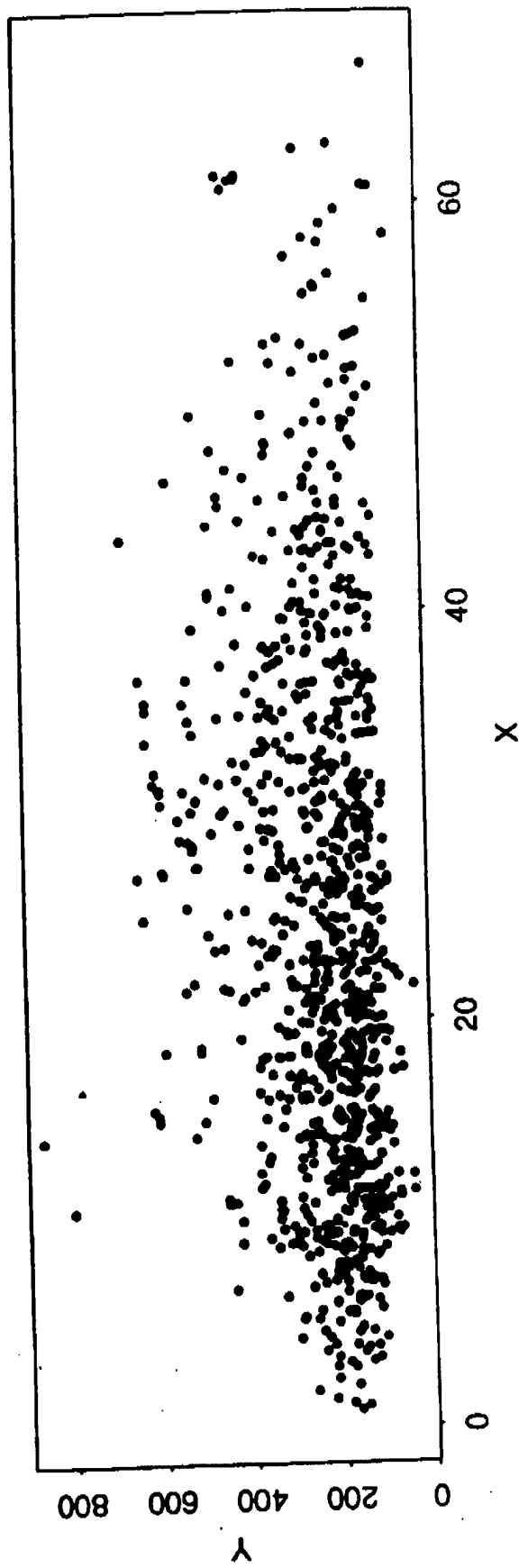


Figure 2a: Armor Penetration Data



241

Figure 2b: 2 Modes Found By Algorithm

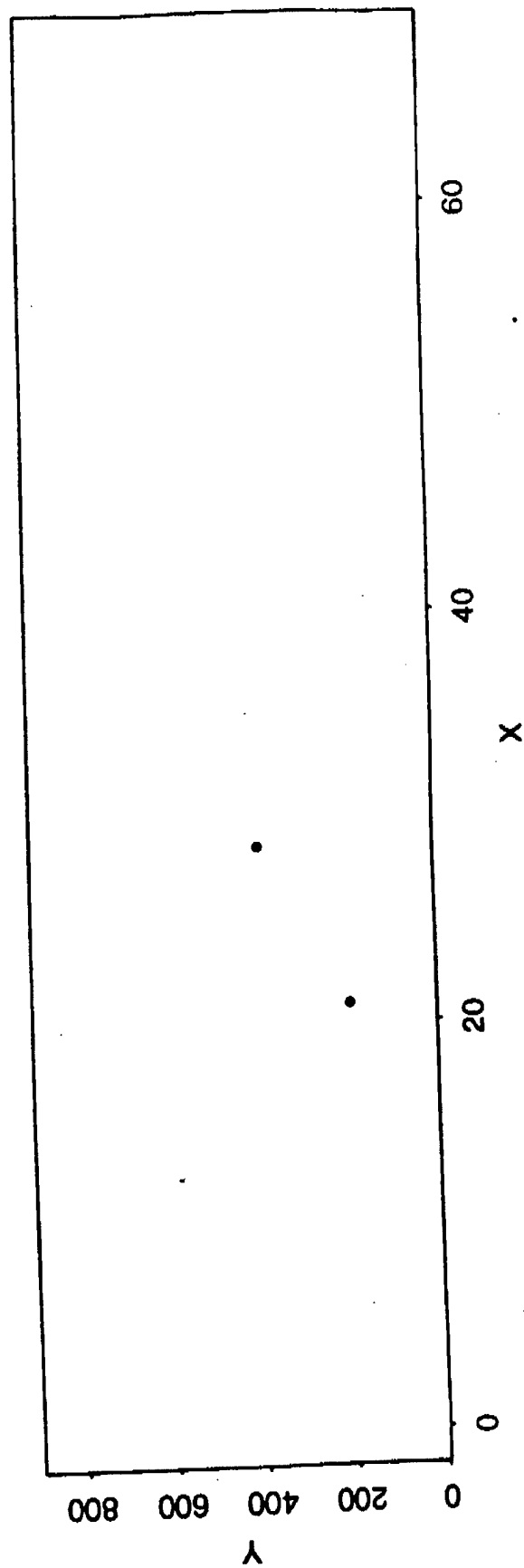


Fig. 3a: Altered Army Data

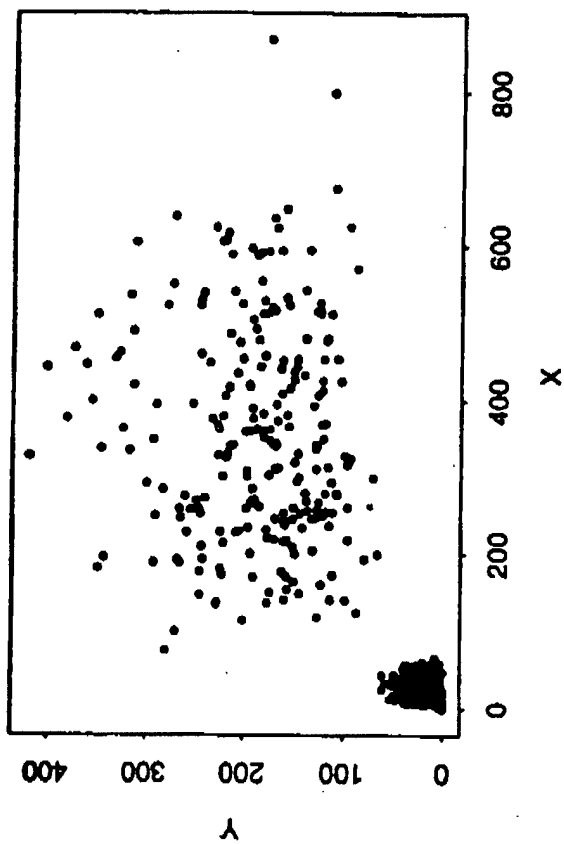


Fig. 3b: 3 Modes Found By Algorithm

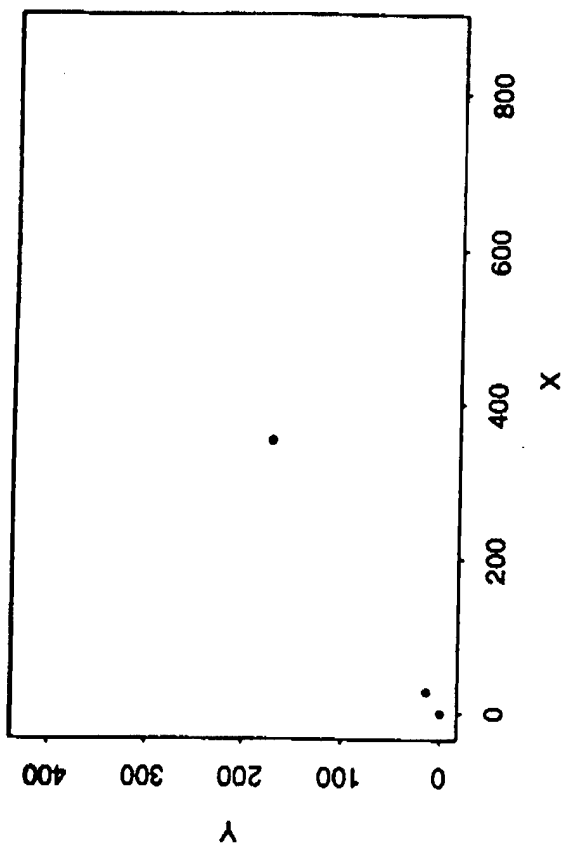


Fig. 3c: Close-Up of Alt. Data

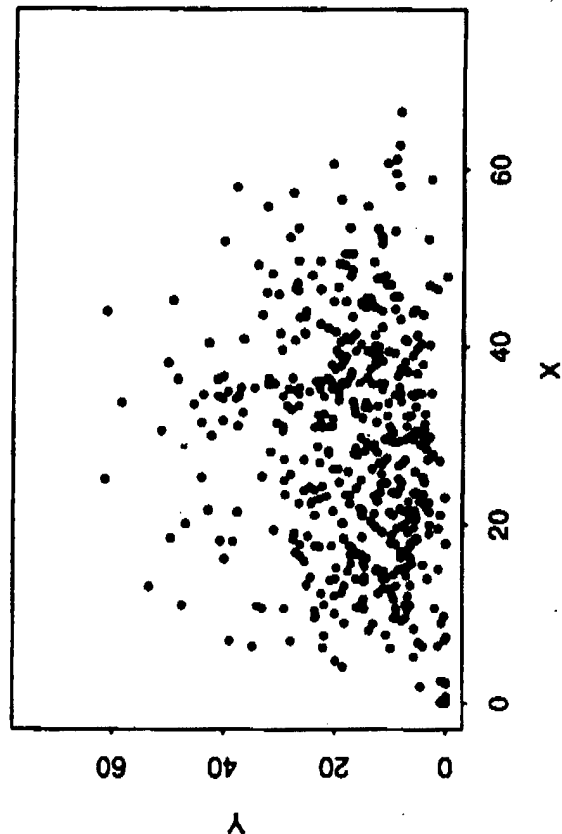


Fig. 3d: Extreme Close-Up of Alt. Data

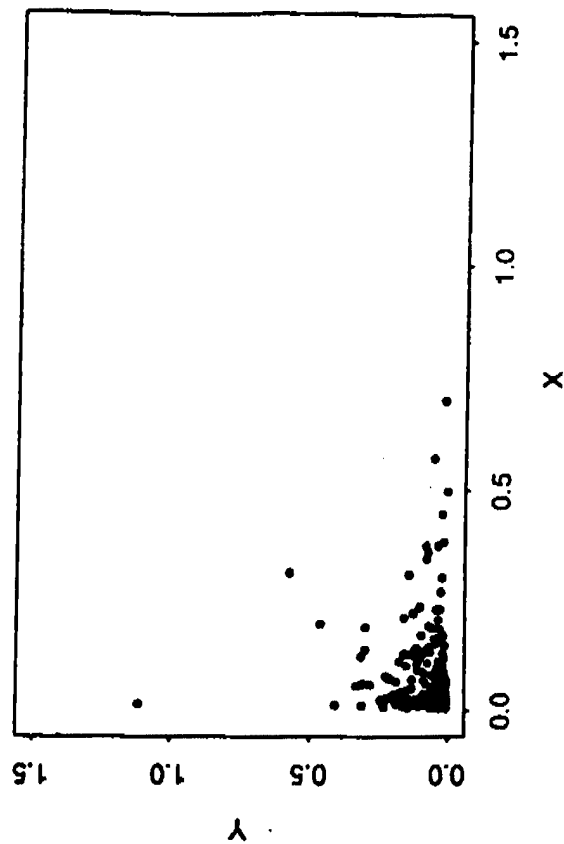




Fig. 4a: Projection of *I. versicolor* and *I. virginica* Data onto Line Connecting Est'd Modes

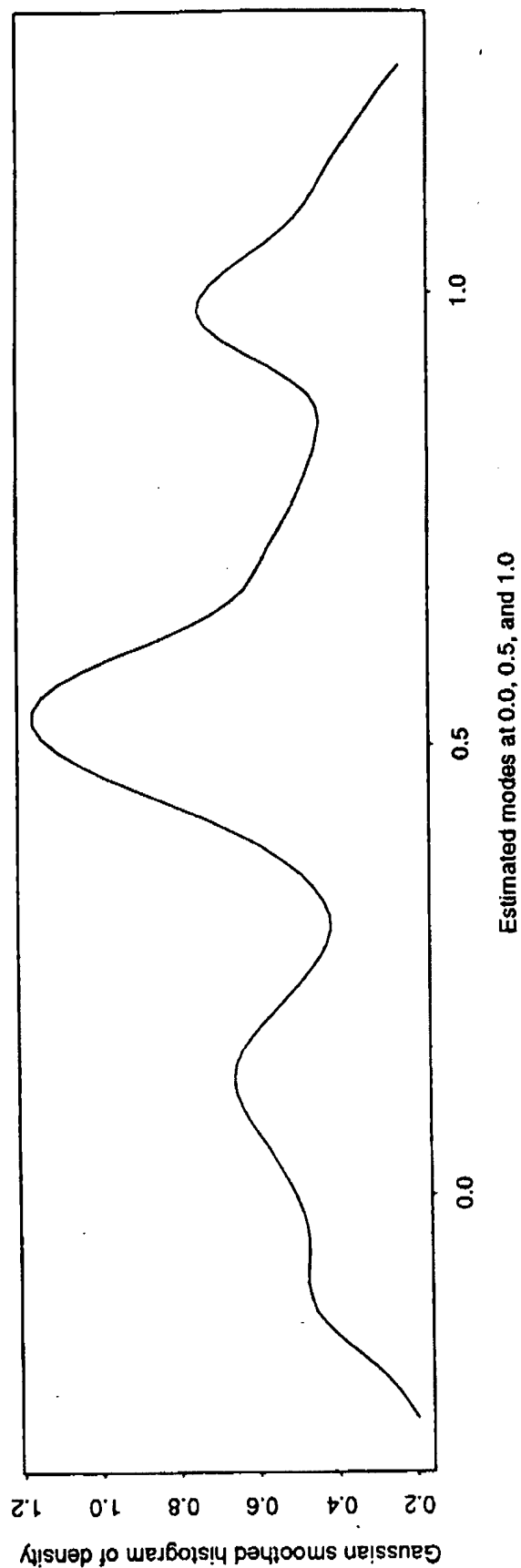
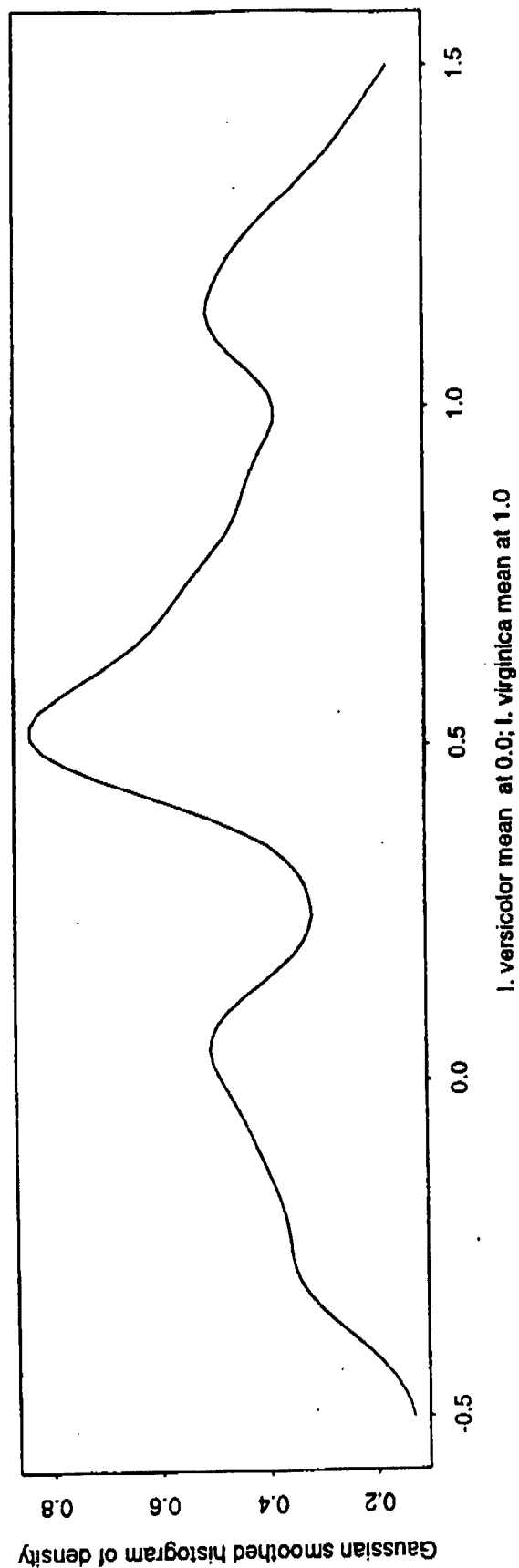


Fig. 4b: Projection of *I. versicolor* and *I. virginica* Data onto Line Connecting Species Means



### Bibliography

- Boswell, S. B. (1983). *Nonparametric Mode Estimation for Higher Dimensional Densities*. Doctoral Dissertation (Rice University).
- Mack, Y. P., & Rosenblatt, (1979). "Multivariate  $k$ -nearest neighbor density estimates," *Journal of Multivariate Analysis*, 9:1-15.
- Silverman, B. W. (1986). *Density Estimation*, New York: Chapman & Hill.
- Taylor, M. S., & Thompson, J. R. (1986). "A data based algorithm for the generation of random vectors," *Computational Statistics and Data Analysis*, v.4, no.2, 93-101.
- Thompson, J. R., & Tapia, R. A. (1990). *Nonparametric Function Estimation, Modelling, and Simulation*. Philadelphia: SIAM.
- Wong & Lane. (1983). *Journal of the Royal Statistical Society Bulletin*, 45: 362-368.

## SAMPLE-WEIGHTED AVERAGE OF VECTORS

Aivars Celmiņš

Advanced Computational and Information Sciences Directorate

U.S. Army Research Laboratory

Aberdeen Proving Ground, MD 21005-5067

**ABSTRACT.** The average burst point of artillery shells is computed from observations of individual burst points in repeated firings. We review such calculations within the framework of general least-squares averaging of observations in an  $n$ -dimensional space and propose to use a "sample-weighted" average in cases where the event scatter (the dispersion of the fire) is unknown and neither it nor the observational errors can be neglected. An iterative computation of the sample-weighted average is presented. The algorithm produces estimates of the vector average and of the event scatter, that is estimates of the expected burst point location and of the dispersion of the artillery fire.

**1. INTRODUCTION.** To determine the accuracy of artillery fire one measures the coordinates of the shell's burst point in repeated firings and calculates an average burst point and its scatter from these measurements. The task amounts to the computation of an average vector in  $R^3$ . The accuracy of each observed vector is known from an analysis of the actual measurements and depends mainly on the geometry of the setup and properties of the measuring instruments. Typically the actual measurements are azimuth and elevation observations from four or more observation towers using theodolites. A non-linear least-squares analysis of these measurements provides for each observed round an estimate of the burst-point coordinate vector together with accuracy estimates of its components. We assume that these accuracy estimates are given in form of an estimated variance-covariance matrix of the components for each observed coordinate vector. If the cannon would fire every time exactly alike (i.e., if the event scatter would be zero) then a reasonable estimate of the burst-point coordinates could be obtained from these vectors by an observation-weighted averaging where the weights are the inverses of the variance-covariance matrices of the observations. However, in practice the event scatter (the dispersion of the artillery fire) can be of the same order or even larger than the measurement scatters and cannot be neglected. Also, in general the principal directions of the event dispersion are different from the principal directions of the measurement-error distributions. Therefore, an observation-weighted averaging in  $R^3$  can have unacceptable results. On

the other hand an unweighted averaging does not take into account the estimates of measurement errors that can be quite different for different rounds. In this paper we define a new "sample-weighted" average of vectors that does not have the disadvantages of observation-weighted or unweighted averages. The definition includes an estimate of the event variance (the dispersion of the artillery fire) that is consistent with the observations and their estimated variances. An iterative algorithm for the computation of the sample-weighted average is suggested.

In Section 2 we define the problem of vector averaging in  $\mathbb{R}^n$  that corresponds to the outlined artillery problem and propose a solution. Section 3 contains some examples and Section 4 is a summary.

**2. ESTIMATION OF AN AVERAGE VECTOR.** Let the observed vectors be  $x_i \in \mathbb{R}^n$ ,  $i = 1, \dots, s$  and let the estimated variance-covariance  $n \times n$  matrices of the observations be  $Q_i$ ,  $i = 1, \dots, s$ . Let the unknown average vector be  $a \in \mathbb{R}^n$  and the variance-covariance matrix of the event be  $P$ . The model equation of the problem is

$$f(x, a) = x - a = 0 \quad (1)$$

We define the least-squares value of  $a$  as the solution of the following constrained optimization problem.

$$\text{Minimize} \quad W = \sum_{i=1}^s (c_i^T Q_i^{-1} c_i + b_i^T P^{-1} b_i) \quad (2)$$

$$\text{subject to} \quad f(x_i + c_i, a + b_i) = x_i + c_i - (a + b_i) = 0, \quad i = 1, \dots, s, \quad (3)$$

where  $c_i$  is the correction of the  $i$ -th observation and  $b_i$  is the deviation of the  $i$ -th event (round) from the average  $a$ .

To solve the minimization problem we introduce a modified objective function  $\tilde{W}$  using Lagrange multiplier vectors  $k_i$ :

$$\tilde{W} = \frac{1}{2} \sum_{i=1}^s (c_i^T Q_i^{-1} c_i + b_i^T P^{-1} b_i) - \sum_{i=1}^s k_i^T (x_i + c_i - a - b_i) \quad (4)$$

We obtain a system of normal equations by setting equal to zero the partial derivatives of  $\tilde{W}$  with respect to  $c_i$ ,  $b_i$ ,  $a$  and  $k_i$ . The result is

$$\left. \begin{aligned} Q_i^{-1}c_i - k_i &= 0, \quad i = 1, \dots, s, \\ P^{-1}b_i + k_i &= 0, \quad i = 1, \dots, s, \\ \sum_{i=1}^s k_i &= 0, \\ x_i + c_i - a - b_i &= 0, \quad i = 1, \dots, s. \end{aligned} \right\} \quad (5)$$

Eliminating the  $k_i$  we obtain the following simpler equation system

$$\left. \begin{aligned} a &= \left[ \sum_{i=1}^s (Q_i + P)^{-1} \right]^{-1} \sum_{i=1}^s (Q_i + P)^{-1} x_i, \\ b_i &= P (Q_i + P)^{-1} (x_i - a), \quad i = 1, \dots, s, \\ c_i &= -Q_i (Q_i + P)^{-1} (x_i - a), \quad i = 1, \dots, s. \end{aligned} \right\} \quad (6)$$

We also obtain

$$W_b = \sum_{i=1}^s b_i^T P^{-1} b_i = \sum_{i=1}^s (x_i - a)^T (Q_i + P)^{-1} P (Q_i + P)^{-1} (x_i - a), \quad (7)$$

$$W_c = \sum_{i=1}^s c_i^T Q_i^{-1} c_i = \sum_{i=1}^s (x_i - a)^T (Q_i + P)^{-1} Q_i (Q_i + P)^{-1} (x_i - a), \quad (8)$$

$$W = W_b + W_c = \sum_{i=1}^s (x_i - a)^T (Q_i + P)^{-1} (x_i - a) \quad (9)$$

and the variance of weight one

$$v_o = \frac{1}{n(s-1)} W. \quad (10)$$

Let the total variance-covariance matrix of the observed  $x_i$  be  $R_{x_i}$  (including both, the measurement scatter and the event scatter). Then the variance-covariance matrix of  $a$  is (see eq. (6))

$$R_a = \left[ \sum_{i=1}^s (Q_i + P)^{-1} \right]^{-1} \sum_{i=1}^s \left[ (Q_i + P)^{-1} R_{x_i} (Q_i + P)^{-1} \right] \left[ \sum_{i=1}^s (Q_i + P)^{-1} \right]^{-1}. \quad (11)$$

If we estimate as usual

$$R_{x_i} = v_o (Q_i + P), \quad i = 1, \dots, s \quad (12)$$

then it follows from eq. (11)

$$R_a = v_o \left[ \sum_{i=1}^s (Q_i + P)^{-1} \right]^{-1} = \left[ \sum_{i=1}^s R_i^{-1} \right]^{-1} . \quad (13)$$

The formulas (6), (9), (10) and (13) provide the general least-squares solution of the averaging problem (defined by eqs. (2) and (3)) if the  $Q_i$  and  $P$  are known. In practice such a situation is an exception, because in general one has estimates of the  $Q_i$  but  $P$  is not known. Therefore, vector averaging is commonly done assuming one of two special cases of the general solution. The special cases are obtained by postulating that either the  $Q_i$  or  $P$  can be neglected in the general solution formulas. We now outline these special cases.

In the first special case one assumes that  $P = 0$ , i.e., that either the event scatter is negligible or that the estimated  $Q_i$  already contain the matrix  $P$ . With this assumption we obtain from eqs. (6) and (13) the usual observation-weighted least-squares averaging formulas:

$$a = \left[ \sum_{i=1}^s Q_i^{-1} \right]^{-1} \sum_{i=1}^s Q_i^{-1} x_i , \quad (14)$$

$$\left. \begin{aligned} b_i &= 0 , & i &= 1, \dots, s , \\ c_i &= a - x_i , & i &= 1, \dots, s , \end{aligned} \right\} \quad (15)$$

$$R_a = v_o \left[ \sum_{i=1}^s Q_i^{-1} \right]^{-1} . \quad (16)$$

Usually the  $Q_i$  are positive definite matrices but in some applications they may be only semi-definite. Also, the sum of their inverses in eqs. (14) and (16) is not necessarily positive definite. The formulas are, however, generally valid if Moore-Penrose generalized inverses are used in both formulas.

In the second case one assumes that the measurement errors are negligible in comparison to the event dispersion, that is,  $Q_i = 0$  for all  $i = 1, \dots, s$  (or that all  $Q_i$  are equal and included in  $P$ ). In that case eqs. (6) and (13) provide the formulas for simple unweighted averaging:

$$a = \frac{1}{s} \sum_{i=1}^s x_i , \quad (17)$$

$$\left. \begin{aligned} b_i &= x_i - a , & i &= 1, \dots, s , \\ c_i &= 0 , & i &= 1, \dots, s , \end{aligned} \right\} \quad (18)$$

$$R_a = v_o \frac{1}{s} P . \quad (19)$$

To complete the calculation in this case we also need an estimate for  $P$ . The usual

estimate is the sample covariance matrix

$$P = \alpha \tilde{P} = \alpha \sum_{i=1}^n b_i b_i^T, \quad (20)$$

where the factor  $\alpha$  is determined such that  $v_o$ , defined by eq. (10), equals unity. Let

$$\tilde{U} = \sum_{i=1}^s (x_i - a)^T \tilde{P}^{-1} (x_i - a). \quad (21)$$

Then the factor is

$$\alpha = \frac{1}{n(s-1)} \tilde{U}. \quad (22)$$

With this value of  $\alpha$  the variance-covariance matrix of the average, eq. (19), becomes

$$R_a = v_o \frac{\alpha}{s} \tilde{P} = \frac{\tilde{U}}{n s (s-1)} \tilde{P} = \frac{1}{n s (s-1)} \sum_{i=1}^s [(x_i - a)^T \tilde{P}^{-1} (x_i - a)] \tilde{P}. \quad (23)$$

As in the first case, one can use the Moore-Penrose generalized inverse in eqs. (21) and (23) if the matrix  $\tilde{P}$  is not positive definite.

In practice, estimates of the  $Q_i$  usually are available but  $P$  is not known so that the general solution formulas (6) cannot be used. If also neither of the two special cases can be justified on basis of additional information then one needs a method to estimate  $P$  from the observations before eqs. (6) can be applied. We propose in such cases to use for  $P$  the estimate (20) through (22). Because this solution makes use of the sample-covariance matrix  $\tilde{P}$  defined in eq. (20), we call the resulting  $a$  the *sample-weighted average*. The numerical computation of the sample-weighted average is complicated by the fact that the unknowns  $b_i$  as well as  $a$  enter eqs. (6) on the right hand sides of the equations also through the definition of  $P$ . We propose the following computing strategy. First, we remove the explicit dependence of  $P$  on  $a$  by seeking a solution for fixed values of the scaling factor  $\alpha$ , that is, by using the estimate (20) with a predetermined  $\alpha$ . We solve this modified problem iteratively. We then vary  $\alpha$  in a second iteration until it satisfies eq. (22). The complete numerical solution process consists of obtaining initial estimates for  $P$  and  $\alpha$ , and an iterative improvement of the initial values.

We initialize the computation with an unweighted averaging

$$a_o = \frac{1}{s} \sum_{i=1}^s x_i, \quad (24)$$

$$b_{o,i} = x_i - a_o, \quad i = 1, \dots, s, \quad (25)$$

and obtain an initial estimate  $P_1$  of  $P$  as follows

$$\tilde{P}_0 = \sum_{i=1}^s b_{0,i} b_{0,i}^T, \quad (26)$$

$$\tilde{U}_0 = \sum_{i=1}^s b_{0,i}^T \tilde{P}_0^{-1} b_{0,i}, \quad (27)$$

$$P_1 = \frac{\tilde{U}_0}{n(s-1)} \tilde{P}_0. \quad (28)$$

Next, we update the initial estimates (24) and (25) of  $a$  and  $b_i$ , respectively, and obtain an initial estimate for the scaling factor  $\alpha$ :

$$a_1 = \left[ \sum_{i=1}^s (Q_i + P_1)^{-1} \right]^{-1} \sum_{i=1}^s (Q_i + P_1)^{-1} x_i, \quad (29)$$

$$b_{1,i} = P_1 (Q_i + P_1)^{-1} (x_i - a_1), \quad i = 1, \dots, s, \quad (30)$$

$$\tilde{P}_1 = \sum_{i=1}^s b_{1,i} b_{1,i}^T, \quad (31)$$

$$\tilde{U}_1 = \sum_{i=1}^s b_{1,i}^T \tilde{P}_1^{-1} b_{1,i}, \quad (32)$$

$$\alpha = \frac{\tilde{U}_1}{n(s-1)}. \quad (33)$$

The actual iteration during which we do update  $P$  but keep the value of  $\alpha$  unchanged is defined by the following iteration formulas for  $k = 1, 2, \dots$

$$P_{k+1} = \alpha \tilde{P}_k, \quad (34)$$

$$a_{k+1} = \left[ \sum_{i=1}^s (Q_i + P_{k+1})^{-1} \right]^{-1} \sum_{i=1}^s (Q_i + P_{k+1})^{-1} x_i, \quad (35)$$

$$b_{k+1,i} = P_{k+1} (Q_i + P_{k+1})^{-1} (x_i - a_{k+1}), \quad i = 1, \dots, s, \quad (36)$$

$$\tilde{P}_{k+1} = \sum_{i=1}^s b_{k+1,i} b_{k+1,i}^T. \quad (37)$$

The variance-covariance estimate  $R_a$  of the average can be computed at each iteration step using eqs. (9), (10) and (13). Iteration end conditions can be expressed, for instance, in terms of changes of the elements of  $a$  and  $R_a$ . Experience shows that the average vector  $a$  becomes stationary after a few iteration steps whereas the elements of  $R_a$  need more steps to meet such convergence criteria. Convergence enhancement techniques were, however, not necessary in numerical experiments with this algorithm.

The numerical result of the iteration depends on the fixed scaling factor  $\alpha$  that was initially set by eq. (33). We want to determine its value such that the variance of



weight one  $v_o$ , defined by eqs. (9) and (10), equals unity. We achieve this by embedding the iteration by eqs. (34) through (37) in a regula falsi algorithm for the solution of the equation  $v_o(\alpha) = 1$ . For this algorithm, we enter the iteration (eqs. (34) through (37)) with the new  $\alpha$ -value and the previous  $\tilde{P}$  as initial estimate, that is, the initializing by eqs. (24) through (33) is used only to obtain a very first approximations of  $\alpha$  and  $\tilde{P}$ . In general, a solution of the equation  $v_o(\alpha) = 1$  with positive  $\alpha$  exists if  $v_o(0) > 1$ , because  $v_o$  decreases with increasing  $\alpha$ . If  $v_o(0) \leq 1$  then the estimated observational errors (the matrices  $Q_i$ ) are so large that the adjustment with  $\alpha = 0$ , i.e., with neglected  $P$  suffices to explain the data. The proper average in such cases is the observation-weighted average defined by eqs. (14) through (16).

The final result of the iterations is the solution (6) and (13) of the general minimization problem, defined by eqs. (2) and (3), whereby the event variance matrix satisfies eq. (20) and  $v_o$  (defined by eq. (10)) equals unity.

**3. EXAMPLES.** We present two examples. The first example is chosen to illustrate the main characteristics of the three types of averaging. In the second example we use actual data.

In the first example we compute the average of three points on a straight line in a plane. The coordinates of the points are (0.5, 2.0), (1.5, 2.1) and (8.5, 2.8). We assume that the observational errors are equal for all three points and given by the following estimate of their variance-covariance matrix

$$Q = \begin{pmatrix} 2.0 & 2.0 \\ 2.0 & 2.0 \end{pmatrix}.$$

The matrix  $Q$  is not positive definite which means that the observational errors are distributed in a subspace of  $\mathbb{R}^2$ , that is, along a straight line. In other words, the observational-error ellipses are degenerated into error bars. Figure 1 shows the data and the observation-weighted average. The coordinates of the average are (3.5, 2.3) and the variance-covariance matrix of the average is

$$R_a = \begin{pmatrix} 0.95792 & 0.95792 \\ 0.95792 & 0.95792 \end{pmatrix}.$$

The corresponding standard-deviation ellipse is again degenerated and shown in Figure 1 as a dashed line. The location of the average point is reasonable but its estimated variance is not because the structure of the variance-covariance matrix that is computed with eq. (16) is independent of the observations and does not reflect the event scatter.

Next we use the same data and compute their unweighted average by eqs. (17) through (23). The average vector is the same as in the previous calculation but its variance-covariance matrix is

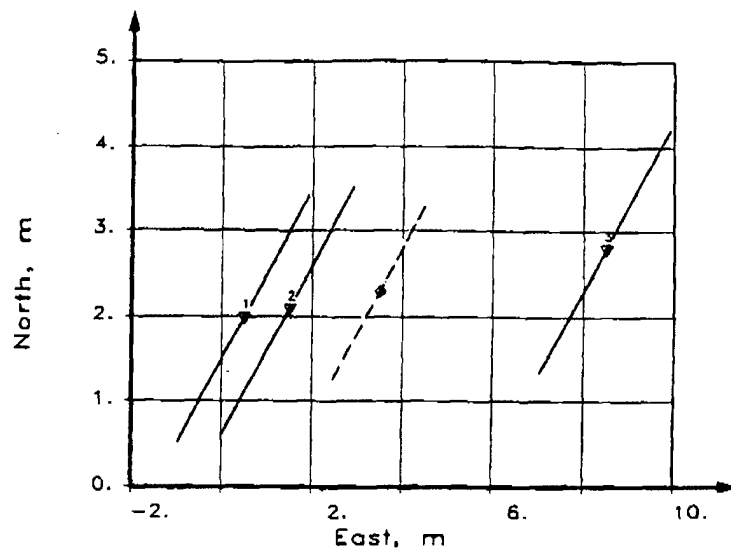


Figure 1. Observation-weighted average.

$$R_a = \begin{pmatrix} 4.75 & 0.475 \\ 0.475 & 0.0475 \end{pmatrix}.$$

The result is shown in Figure 2. The image of the one-standard-error ellipse of the average is an error bar in the direction of the scatter of the observations, because in this case  $R_a$  is independent of the observational-error variances.

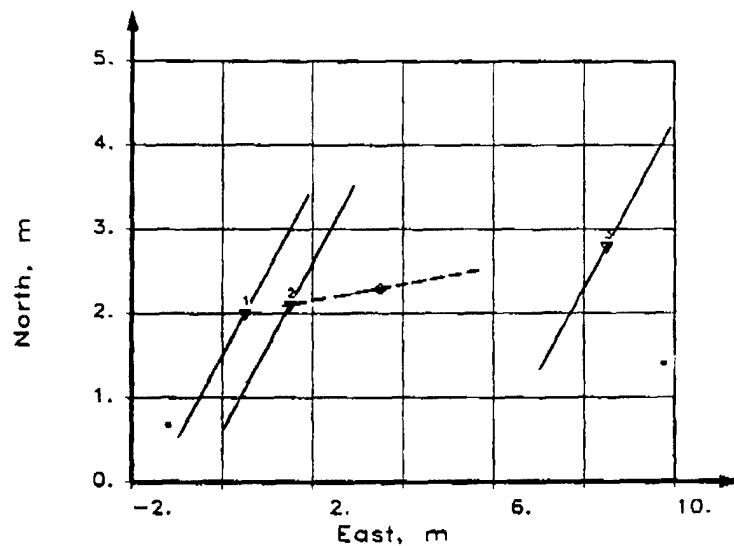
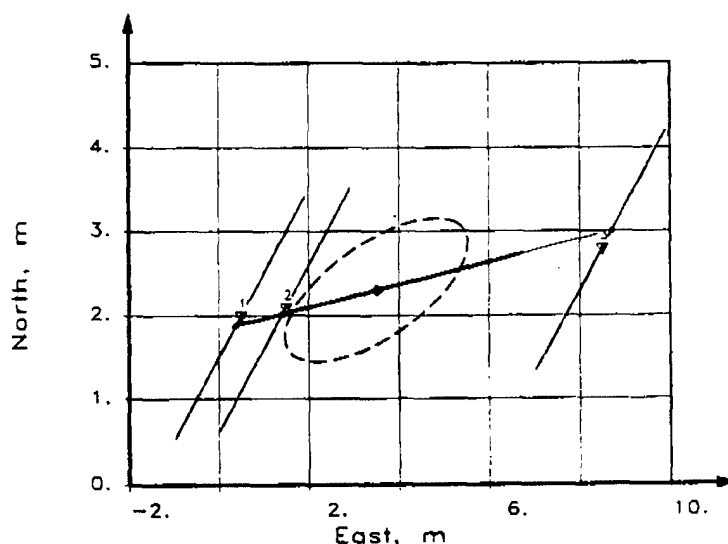


Figure 2. Unweighted average.

Finally, we compute the sample-weighted average. The average vector again is the same as before. Its variance-covariance matrix is

$$R_a = \begin{pmatrix} 4.12682 & 1.13567 \\ 1.13567 & 0.73027 \end{pmatrix}.$$

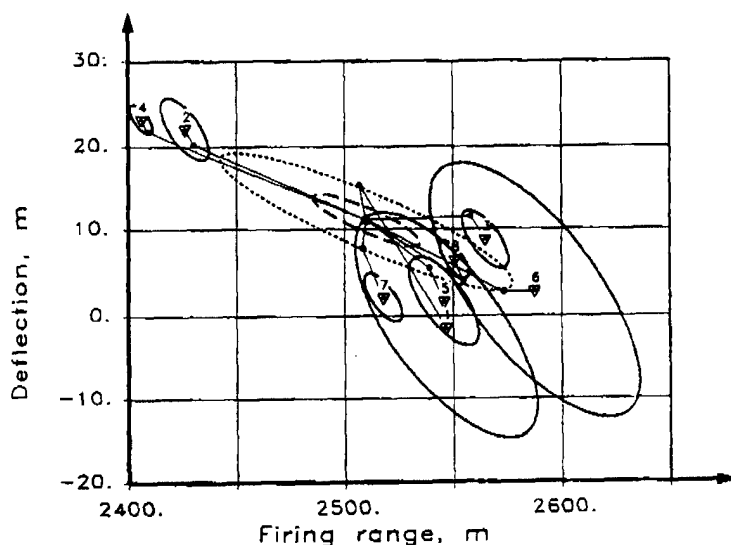
Figure 3 shows the corresponding one-standard-error ellipse. The figure also contains the correction vectors  $b_i$  plotted as rays from the average point. The end points of the  $b_i$  are indicated by dots. In this example, all vectors  $b_i$  are parallel so that their end points are located on a straight line and the matrix  $P$ , eq. (34), is only positive semi-definite. The image of the ellipse representing  $P$  is the heavier segment of the straight line in the direction of the  $b_i$ . The differences between the dots and the corresponding observations (inverted triangles) are the corrections  $c_i$ . We observe that all corrections  $c_i$  and  $b_i$  are in the direction of the corresponding error bars, as they should be. In this example the iteration with eqs. (34) through (37) became stationary after two steps. The initial scaling factor and the variance of weight one were, respectively,  $\alpha = 0.250$  and  $v_o = 1.008$ . After three regula falsi steps, we had the values  $\alpha = 0.252006$  and  $v_o = 1.00006$ .



**Figure 3. Sample-weighted average.**

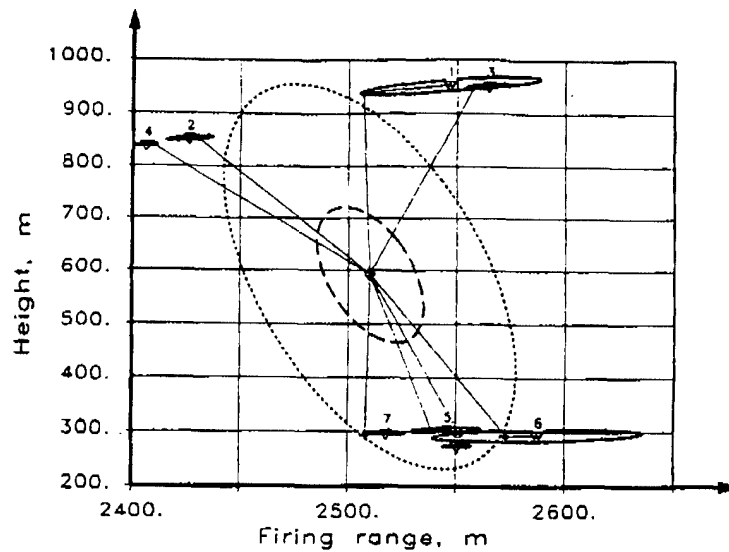
In our second example we use actual observations of artillery burst-point coordinates. The observations  $x_i$  are three-dimensional vectors that specify the range, deflection and height of the burst. The vectors were obtained from simultaneous measurements of directional angles (azimuths and elevations) of the burst points from four observation towers. A least-squares reduction of the eight measurements of each observed round provided the three components of the burst location vector  $x_i$  and an

estimate of its accuracy in form of the variance-covariance matrix  $Q_i$ . The observation set in our example consisted of eight observed burst points from the same howitzer. The estimated accuracies of the observations varied widely between rounds and were smaller than the scatter between the burst points, but not negligible. Figures 4, 5 and 6 show the observed points as inverted triangles and their sample-weighted average as a diamond. The figures also contain the projections of the one-standard-error ellipsoids corresponding to the estimated  $Q_i$ . The standard-error ellipsoid of the average, defined by  $R_a$ , is plotted with a dashed line. The standard-deviation ellipsoid of a single shot, i.e., the dispersion of the artillery fire is defined by the matrix  $P$  and plotted with a dotted line. We note that contrary to the appearance in the plots  $P$  is not proportional to  $R_a$ : the relation between  $P$  and  $R_a$  is given by eq. (13). The correction vectors  $b_i$  represent the deviation of the round  $i$  from the average and are plotted as rays from the average point, as in Figure 3. We observe that these corrections in general do not point in the directions towards the observations  $x_i$ , but in other directions such that both corrections,  $b_i$  and  $c_i$ , are in directions of large error estimates thus minimizing  $W$ , eq. (2). The initial estimate of the scaling factor was  $\alpha = 0.143$  and the variance of weight one was  $v_o = 1.137$ . After four regula falsi steps, the results were  $\alpha = 0.176998$  and  $v_o = 1.00004$ . The iteration for  $a$  and  $b_i$ , eqs. (34) through (37), required eight steps at the beginning and three steps at the end of the regula falsi calculations.

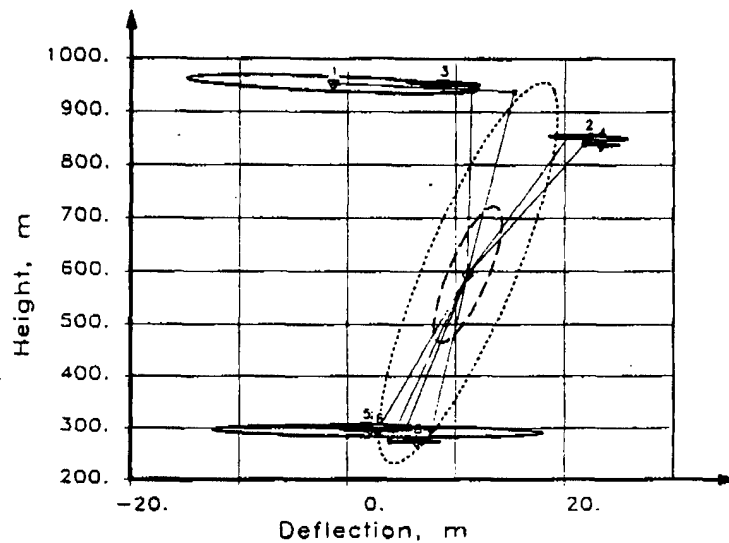


**Figure 4. Burst-point range and deflection.**

To illustrate the advantage of the sample-weighted average we show in Figures 7, 8 and 9 the usual observation-weighted average [eqs. (14) through (16)] of the same observations. We notice that in this example the observation-weighted average is



**Figure 5. Burst-point range and height.**



**Figure 6. Burst-point deflection and height.**

completely useless: the estimated burst point is shifted far outside the cloud of observations. From an inspection of the figures, we conclude that this shift is caused by the high sensitivity of the location of the average to the estimated principal directions of observational errors. The variance of weight one was in this case  $v_0 = 5996$  indicating that measurement errors alone are not sufficient to explain the data scatter.

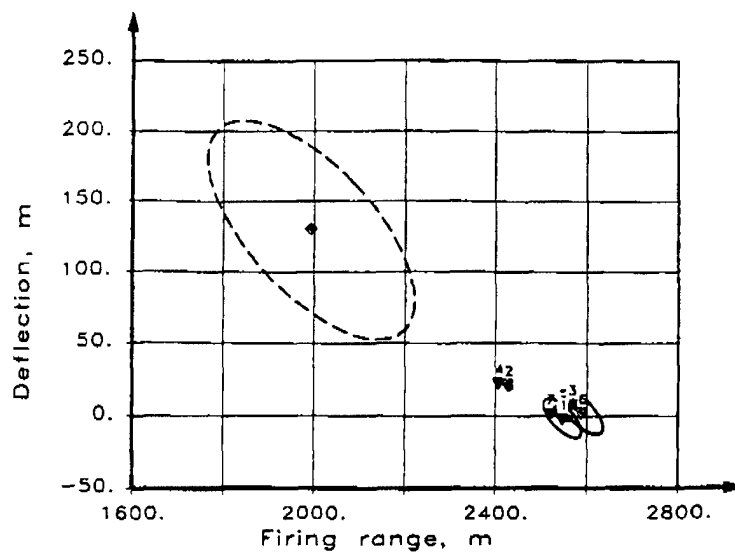


Figure 7. Observation-weighted range and deflection.

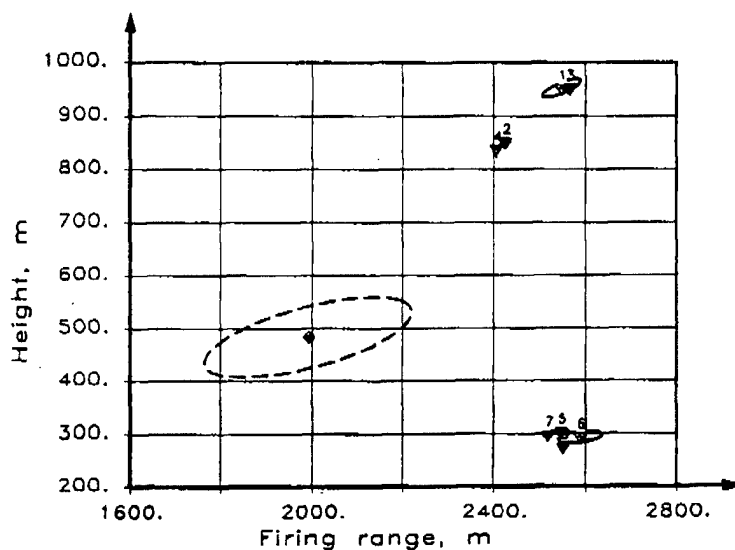
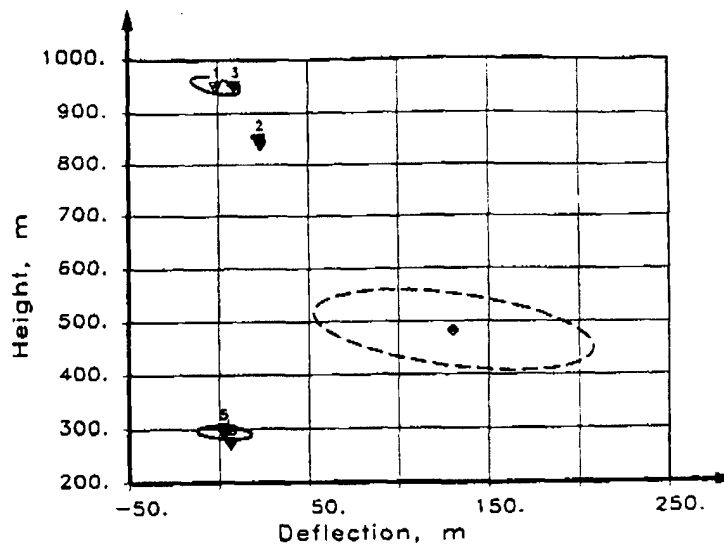


Figure 8. Observation-weighted range and height.

**4. SUMMARY.** We have considered least-squares computations of vector averages. We assume that the observations in a  $n$ -dimensional space contain inaccuracies from two sources: observational errors and variations of the observable itself, that is, event scatter. Usually one of these error sources is neglected. If a simple unweighted average is computed then one assumes implicitly that the observational errors are negligible. If an observation-weighted average is computed then the implicit



**Figure 9. Observation-weighted deflection and height.**

assumption is that the event scatter can be neglected. Most often the event variation is not known and one has no grounds for using one of these special averages. If event scatter is known to exist we propose to use the sample-weighted average. It can be computed by an iterative algorithm that provides in addition to the average of the observed vectors with its variance, also an estimate of the variance-covariance matrix of the event.

Applied to the computation of average burst points of artillery fire the sample-weighted averaging provides a burst-point estimate that is consistent with observations and their error estimates. It also produces a consistent estimate of the dispersion of the fire.

## JUMP CHARACTERIZATION TEST

CHARLES E. HEATWOLE  
RELIABILITY, AVAILABILITY AND MAINTAINABILITY DIVISION  
U.S. ARMY MATERIEL SYSTEMS ANALYSIS ACTIVITY  
ABERDEEN PROVING GROUND, MD 21005-5071

ABSTRACT. Tank jump is a little-understood - but major - element of tank gun accuracy. Current tank gunnery doctrine utilizes a computer correction factor for the tank fire control computer to correct for the mean jump of each tank munition.

Numerous accuracy tests of U.S. tank munitions have repeatedly indicated that several factors have a highly significant effect on jump. However, no testing has been done to characterize the occasion-to-occasion variation in jump when these factors are held constant. The U.S. Army Materiel Systems Analysis Activity (AMSAA) has proposed such a test. The goal of this test is to improve our understanding of jump in an attempt to determine if more effective correction factors can be developed.

1. INTRODUCTION. Identification of the various factors affecting weapon system accuracy has been a long-standing problem. Ideally, the primary factors can be identified and either compensated for or eliminated. For tank munitions, the fire control computer system corrects for certain known variables such as wind, temperature, etc. The remaining error not specifically accounted for is referred to as jump. Until approximately 1981, individual tank zeroing was performed as a means for compensating for this remaining error. However, since then, the fleet zero concept has been adopted.

Under the fleet zero concept, instead of an individual correction factor being determined for each tank, a common correction factor is used by all tanks for a given ammunition type. This standardized correction factor is referred to as the "Computer Correction Factor" (CCF). To determine a CCF for each munition, accuracy testing is performed under a variety of tank firing conditions. For each round fired, jump is calculated from the data collected. Essentially, the mean jump is then used as the CCF. This value is used by every tank in the fleet as a final add-on correction to the fire control computer's ballistic solution computed immediately prior to firing.

The utilization of the CCF/fleet zero concept is predicated on the assumption that jump is a fixed bias (i.e., is consistent in magnitude from one occasion to another). However, analyses of variance performed on jump data for various tank munitions fired over the past few years have repeatedly indicated that gun tube and ammunition temperature (even after application of the fire control correction for temperature) are highly significant influences on jump, and that several other factors sometimes have a significant effect. Hence, statistically, jump is not a fixed bias. This has prompted the question as to what extent jump varies from occasion-to-occasion without any intentional tank or temperature-related changes. That is, how much variation in jump occurs from occasion to occasion when the only intended change is disassembly of the test setup upon completion of an occasion, and re-setup to start a new occasion?



It is planned that there be approximately two weeks between replications, with the primary difference between replications being disassembly of the test setup after completion of a given replication, and re-setup upon starting the next replication. It is also intended that the two tanks used in the test be dedicated solely to this test until completion; i.e., there should be no other firings or mileage accrued beyond the mileage required for normal travel to and from the weapon shops, to minimize the chance of extraneous influences affecting the tanks between replications.

3. ANALYSIS. Of the three factors, the effect of replication on jump is the greatest concern, and the effect of ammunition lot is the next factor of interest. There has been extensive discussion as to whether replication, lot, and tank should be considered fixed or random effects. In all cases, it is desired that inferences can be made regarding the population. For example, the estimate of replication effect would be considered representative of the general occasion-to-occasion variability in jump within constant firing conditions, which implies that replication should be treated as a random factor. The same is true for tank and lot number. However, each tank (and its tube) is being selected from those available at Aberdeen Proving Ground, and not randomly from the worldwide fleet. The lots will be selected from a group of approximately thirty lots available within the U.S. (although they have been previously stored under various conditions overseas). Thus, tank/tube and lot number do not seem to be either purely random or purely fixed by strict definition. It is currently planned to treat these as random factors, because of the desire to draw inferences to the population. However, it is also possible that the analysis should be performed both ways (i.e., treating them as fixed, and treating them as random) to determine if the results are significantly affected.

The breakdown of degrees of freedom for this experiment is:

<u>FACTOR</u>	<u>DEGREES OF FREEDOM</u>
Tank (T)	1
Lot (L)	4
Replication (R)	3
TxL	4
TxR	3
LxR	12
TxLxR	12
<u>Error</u>	<u>80</u>
TOTAL	119

This breakdown is based on analysis of the experiment as a randomized block design (i.e., disregarding the incomplete randomization within replications and the associated possibility of a day effect being confounded with tank effect). If there is concern about the possibility of a day effect within replications, the results for tanks A and B could each be analyzed separately using the following degrees of freedom breakdown:

<u>FACTOR</u>	<u>DEGREES OF FREEDOM</u>
Lot (L)	4
Replication (R)	3
LxR	12
<u>Error</u>	<u>40</u>
TOTAL	59

A test has been designed by the U.S. Army Materiel Systems Analysis Activity (AMSAA) to address this question. The objective of this clinical presentation is to describe the proposed test and solicit advice regarding the design and the associated analyses.

2. TEST DESCRIPTION. The following test matrix has been proposed for this test:

#### JUMP CHARACTERIZATION TEST MATRIX

<u>Ammo Lot</u>	OCCASION							
	I		II		III		IV	
	<u>Tank</u>		<u>Tank</u>		<u>Tank</u>		<u>Tank</u>	
	A	B	A	B	A	B	A	B
1	3	3	3	3	3	3	3	3
2	3	3	3	3	3	3	3	3
3	3	3	3	3	3	3	3	3
4	3	3	3	3	3	3	3	3
5	3	3	3	3	3	3	3	3

This experiment can be described as either a two-factor experiment with replication, or a three-factor experiment with replication as one of the factors. The two basic factors are ammunition lot number and tank, with five lots and two tanks planned for the test. Each lot will be tested with each tank (3 rounds fired per lot/tank/replicate). This factorial scheme is replicated four times. Ideally, the experiment would be conducted in a randomized block fashion, with the order of the lot/tank combinations randomized within each replicate. However, because of the extensive setup time required whenever a different tank is to be fired at the specially-instrumented test range used for tank accuracy testing, it is strongly preferred that all 15 rounds scheduled for a given tank in a given replication be fired once that tank is set up. Thus, the test groups from tank A would not be co-mingled with those from tank B. Randomization within a replication would be limited to randomizing the firing order of the fifteen test rounds within each tank. Once a tank is set up for firing, the fifteen rounds would then be fired within approximately 2 1/2 to 3 hours, with the stipulation that no intentional breaks be allowed in the firing cadence beyond the usual amount of time required to prepare the next round for firing. The setup and firing time for one tank (within a replication) would require the majority of a work day. Therefore, two days will be required to complete each replication, with tank A fired on one day and tank B on the other. It is intended that these days be consecutive to minimize the possibility of day-to-day influences. The order of firing for tanks will be randomized from one replicate to another (e.g., Tank A can be fired on Day 1 in some replicates and Day 2 in others).

The sample size of three rounds per replication was dictated by the availability of no more than 28 rounds per lot for this test. The proposed test matrix would utilize 24 of the 28 rounds, leaving four per lot for use as replacement rounds to be fired if data are lost for any of the test rounds, or if it is necessary to repeat an entire day's firings to achieve two consecutive firing days. Previous tank accuracy test data have indicated that three-round groups are generally sufficient, and a smaller group size is undesirable.

For either method of analysis, plots of the data will be prepared to allow for quick checks of homoscedasticity and evidence of trends.

If the analysis of variance of the jump data indicates the occurrence of any significant influences, similar analyses of variance will be performed on the lower-level data elements used to compute jump. These additional analyses will be performed to help identify the significant components of jump.

4. QUESTIONS FOR DISCUSSION. The following questions are requested as discussion topics relative to the proposed test design and analyses:

a. Given the ammunition hardware limitations of five lots, 28 rounds per lot, and the daily firing time constraints, is the proposed design a reasonable approach for investigating occasion-to-occasion jump variation? Is there a better (more informative or more conclusive) approach?

b. Given that this design and associated firing procedure are used, is analysis of it as a randomized block design affected because of the incomplete randomization within replications?

c. Should the analysis be approached in a different manner due to the possibility that a day effect could be confounded with a tank effect?

d. Is it appropriate to treat the factors as random factors?

# Easy-to-Apply Results for Establishing Convergence of Markov Chains in Bayesian Analysis

Krishna B. Athreya<sup>1</sup>

*Department of Statistics  
Iowa State University  
Ames, Iowa 50011*

Hani Doss<sup>2</sup> and Jayaram Sethuraman<sup>3</sup>

*Department of Statistics  
Florida State University  
Tallahassee, Florida 32306-3033*

February 1993

FSU Technical Report No. 884  
AFOSR Series D Technical Report No. 10  
USARO Technical Report No. 133

---

<sup>1</sup> Research supported by National Science Foundation Grant DMS-92-04938.

<sup>2</sup> Research supported by Air Force Office of Scientific Research Grant 90-0202.

<sup>3</sup> Research supported by Army Research Office Grant DAAL03-90-G-0103.

# Easy-to-Apply Results for Establishing Convergence of Markov Chains in Bayesian Analysis

Krishna B. Athreya \*

Iowa State University

Hani Doss † and Jayaram Sethuraman ‡

Florida State University

## Abstract

The Markov chain simulation method has become a powerful computational method in Bayesian analysis. The success of this method depends on the convergence of the Markov chain to its stationary distribution. We give two carefully stated theorems, whose conditions are easy to verify, that establish this convergence. We give versions of our conditions which are simpler to verify for the Markov chains that arise most commonly in Bayesian analysis.

*Key words and phrases:* Bayesian Poisson regression; calculation of posterior distributions; ergodic theorem; Markov chain simulation method.

## 1 Introduction

Let  $\pi$  be a probability distribution on a measurable space  $(\mathcal{X}, \mathcal{B})$ . The Monte Carlo Markov chain method is a technique for estimating characteristics of  $\pi$  such as  $\pi(E)$  or  $\int f d\pi$  where  $E \in \mathcal{B}$  and  $f$  is a bounded measurable function, and which is useful when  $\pi$  is too complex to describe analytically. The idea is very

---

\*Research supported by National Science Foundation Grant DMS-92-04938

†Research supported by Air Force Office of Scientific Research Grant 90-0202

‡Research supported by Army Research Office Grant DAAL03-90-G-0103

straightforward. We construct a transition probability function  $P(x, \cdot)$  with the property that it has stationary distribution  $\pi$ , i.e.

$$\pi(C) = \int P(x, C)\pi(dx) \text{ for all } C \in \mathcal{B}. \quad (1.1)$$

Then, we generate a Markov chain  $\{X_n\}$  with this transition probability function as follows. We fix a starting point  $x_0$ , generate an observation  $X_1$  from  $P(x_0, \cdot)$ , generate an observation  $X_2$  from  $P(X_1, \cdot)$ , etc. This produces the Markov chain  $x_0 = X_0, X_1, X_2, \dots$ . We use this construction in one of two ways. Either we discard an initial segment  $X_0, X_1, X_2, \dots, X_r$  of the Markov chain, in which the chain has not yet converged to its stationary distribution, and retain the rest of the chain, or we independently run a large number of chains and for each retain only the last observation. In either case we use the observations that we have kept to obtain empirical estimates of those features of  $\pi$  that are of interest.

Implicit in this method is the assumption that the chain converges to its stationary distribution, for a wide class of starting points  $x_0$ . Indeed, one can easily give examples of Markov chains that do not converge to their stationary distribution from any starting point. Thus, to establish the validity of the method, it is crucial to obtain results that give conditions which imply convergence of the chain.

The Markov chain literature already contains many results that give conditions under which the Markov chain converges to its stationary distribution for a class of starting points  $x_0$  which have probability one under  $\pi$  (this condition is called *ergodicity*). Unfortunately, when one comes to apply these results, one immediately notices that in *statistical* applications, the conditions of these theorems are virtually impossible to check.

In our work we have obtained two theorems (Theorems 1 and 2 below) that assert ergodicity of the chain under conditions that are extremely easy to verify in a wide range of problems that are likely to arise in Bayesian statistics. These theorems pertain, roughly, to the two modes of using the Markov chain construction. Before explaining our theorems, it is useful to give an idea of the wide scope of the problems that can be approached via the Monte Carlo Markov chain method.

There are many ways to produce a transition function satisfying (1.1). Methods include the Metropolis algorithm and its variants, and the so-called Gibbs sampler. This last method appears to be the one that is the most widely used in Bayesian statistics, and we now proceed to describe it. This algorithm is used to estimate the unknown joint distribution  $\pi = \pi_{X^{(1)}, \dots, X^{(p)}}$  of the (possibly vector-valued) random variables  $(X^{(1)}, \dots, X^{(p)})$  by updating the coordinates one at a time, as follows. We suppose that we know the conditional distributions  $\pi_{X^{(i)}|\{X^{(j)}_{j \neq i}\}}$ ,  $i = 1, \dots, p$  or at least that we are able to generate observations from these conditional distributions. If  $X_m = (X_m^{(1)}, \dots, X_m^{(p)})$  is the current state, the next state  $X_{m+1} = (X_{m+1}^{(1)}, \dots, X_{m+1}^{(p)})$  of the Markov chain is formed as follows. Generate  $X_{m+1}^{(1)}$  from  $\pi_{X^{(1)}|\{X^{(j)}_{j \neq 1}\}}(\cdot, X_m^{(2)}, \dots, X_m^{(p)})$ , then  $X_{m+1}^{(2)}$  from

$\pi_{X^{(2)}|\{X^{(j)}\}_{j \neq 2}}(X_{(m+1)}^{(1)}, \cdot, X_m^{(3)}, \dots, X_m^{(p)})$ , and so on until  $X_{m+1}^{(p)}$  is generated from  $\pi_{X^{(p)}|\{X^{(j)}\}_{j \neq p}}(X_{(m+1)}^{(1)}, \dots, X_{(m+1)}^{(p-1)}, \cdot)$ . If  $P$  is the transition function that produces  $X_{m+1}$  from  $X_m$ , then it is easy to see that  $P$  satisfies (1.1).

We now give a very brief description of how this method is useful in some Bayesian problems. We suppose that the parameter  $\theta$  has some prior distribution, that we observe a data point  $Y$  whose conditional distribution given  $\theta$  is  $\mathcal{L}(Y|\theta)$ , and that we wish to obtain  $\mathcal{L}(\theta|Y)$ , the conditional distribution of  $\theta$  given  $Y$ . It is often the case that if we consider an (unobservable) auxiliary random variable  $Z$ , then the distribution  $\pi_{\theta,Z} = \mathcal{L}(\theta, Z|Y)$  has the property that  $\pi_{\theta|Z} (= \mathcal{L}(\theta|Y, Z))$  and  $\pi_{Z|\theta} (= \mathcal{L}(Z|Y, \theta))$  are easy to calculate. Typical examples are missing and censored data problems. If we have a conjugate family of prior distributions on  $\theta$ , then we may take  $Z$  to be the missing or the censored observations, so that  $\pi_{\theta|Z}$  is easy to calculate. The Gibbs sampler then gives a random observation with distribution (approximately)  $\mathcal{L}(\theta, Z|Y)$ , and retaining the first coordinate gives an observation with distribution (approximately) equal to  $\mathcal{L}(\theta|Y)$ .

Another application arises when the parameter  $\theta$  is high dimensional, and we are in a nonconjugate situation. Let us write  $\theta = (\theta_1, \dots, \theta_k)$ , so that what we wish to obtain is  $\pi_{\theta_1, \dots, \theta_k}$ . Direct calculation of the posterior will involve the evaluation of a  $k$ -dimensional integral, which may be difficult to accomplish. On the other hand, application of the Gibbs sampler involves the generation of one-dimensional random variables from  $\pi_{\theta_i|\{\theta_j\}_{j \neq i}}$ . The generation of random variables from a one-dimensional distribution is in general much easier than from a multidimensional distribution; very often special tricks can be used. We illustrate this with an example in Section 2 below. In addition, we note that the distribution  $\pi_{\theta_i|\{\theta_j\}_{j \neq i}}$  is available in closed form, except for a normalizing constant. There exist very efficient algorithms for generating random variables from such a distribution, provided the distribution is unimodal; see Zaman (1992).

## 2 Illustration of the Markov Chain Simulation Method: Bayesian Poisson Regression

As a typical application of how the Gibbs sampler helps in high dimensional problems, we consider a model involving Bayesian Poisson regression. This model is

$$Y_i \sim \text{Poisson}(\lambda_i), \quad \lambda_i = \sum_{j=1}^p x_{ij}\beta_j, \quad i = 1, 2, \dots, n,$$

where the  $x_{ij}$ 's are non-negative covariates, and where the prior distribution on the  $\beta_j$ 's is a product of Gammas. In this case, the likelihood function is

$$p(\lambda) = \prod_{i=1}^n \exp(-\lambda_i) \frac{\lambda_i^{y_i}}{y_i!}$$

$$= (y_1! y_2! \dots y_n!)^{-1} \exp \left( - \sum_{i=1}^n \sum_{j=1}^p x_{ij} \beta_j \right) \prod_{i=1}^n \left( \sum_{j=1}^p x_{ij} \beta_j \right)^{y_i}$$

and the joint density of the  $\beta_j$ 's is given by

$$\begin{aligned} f_{\beta}(\beta_1, \beta_2, \dots, \beta_p) &= \prod_{j=1}^p \frac{b_j^{a_j}}{\Gamma(a_j)} \beta_j^{a_j-1} \exp(-b_j \beta_j) \\ &\propto \exp \left( - \sum_{j=1}^p b_j \beta_j \right) \prod_{j=1}^p \beta_j^{a_j-1}, \end{aligned}$$

where  $a_j$  is the shape and  $b_j$  is the scale parameter for the distribution of  $\beta_j$ ,  $j = 1, 2, \dots, p$ . The posterior joint density of the  $\beta_j$ 's, given the data, is therefore

$$\pi(\beta_1, \beta_2, \dots, \beta_p) \propto \exp \left( - \sum_{j=1}^p \beta_j v_j \right) \prod_{j=1}^p \beta_j^{a_j-1} \left( \prod_{i=1}^n \left( \sum_{j=1}^p x_{ij} \beta_j \right)^{y_i} \right),$$

where  $v_j = b_j + \sum_{i=1}^n x_{ij}$ ,  $j = 1, 2, \dots, p$ . To determine the posterior joint density of the  $\beta_j$ 's exactly, the constant of proportionality needs to be determined. This requires high-dimensional integration. However, the Gibbs sampler can be used if we know the conditional distributions of any  $\beta_i$  given the rest of the  $\beta_j$ 's and the data.

To compute the conditional density of any  $\beta_k$ ,  $k = 1, 2, \dots, p$ , given the rest of the  $\beta_j$ 's and the data, we proceed as follow. For each  $l$ ,  $1 \leq l \leq p$ , let  $S_l = \{1, 2, \dots, p\} \setminus \{l\}$ . Then for each  $k$ , the density of  $\beta_k$ , conditional on all  $\beta_j$ ,  $j \in S_k$ , and the data is the discrete mixture of Gamma densities

$$f_{\beta_k | \beta_j, j \in S_k}(\beta_k) \propto \beta_k^{a_k-1} \exp(-v_k \beta_k) \prod_{i=1}^n (c_i + x_{ik} \beta_k)^{y_i},$$

where  $c_i = \sum_{j \in S_k} x_{ij} \beta_j$ . Let  $m = \sum_{i=1}^n y_i$  and write

$$\prod_{i=1}^n (c_i + x_{ik} \beta_k)^{y_i} = \sum_{l=0}^m r_l(k) \beta_k^l,$$

where we explicitly show the dependence of the coefficients on  $k$ . Then,

$$f_{\beta_k | \beta_j, j \in S_k}(\beta_k) \propto \sum_{l=0}^m r_l(k) \beta_k^{a_k+l-1} \exp(-v_k \beta_k)$$

and we readily recognize that

$$f_{\beta_k | \beta_j, j \in S_k}(\beta_k) = \sum_{l=0}^m p_l(k) g_{a_k+l, v_k}(\beta_k),$$

where  $g_{p,q}(x)$  denotes the gamma density with shape parameter  $p$  and scale parameter  $q$  in  $x$ , and  $p_l(k) = r_l(k) \Gamma(a_k + l) / v_k^{a_k+l}$ . The  $p_l(k)$ 's are the discrete mixture probabilities.



### 3 Convergence Theorems

Before stating our theorems, we will need a few definitions concerning Markov chains. Let  $P^n(x, \cdot)$  denote the distribution of  $X_n$  when the chain is started at  $x$ . Also, for a set  $C \in \mathcal{B}$ , let  $T(C) = \inf\{n : n > 0, X_n \in C\}$  be the first time the chain hits  $C$ , after time 0. Finally, for any subset  $\mathcal{I}$  of the positive integers,  $\text{g.c.d.}(\mathcal{I})$  will denote the greatest common divisor of the integers in  $\mathcal{I}$ .

**Theorem 1** Suppose that the Markov chain  $\{X_n\}$  with transition function  $P(x, C)$  has an invariant probability measure  $\pi$ , i.e. (1.1) holds. Suppose that there is a set  $A \in \mathcal{B}$ , a probability measure  $\rho$  with  $\rho(A) = 1$ , a constant  $\epsilon > 0$ , and an integer  $n_0 \geq 1$  such that

$$\pi\{x : P_x(T(A) < \infty) > 0\} = 1, \quad (3.1)$$

and

$$P^{n_0}(x, \cdot) \geq \epsilon \rho(\cdot) \text{ for each } x \in A. \quad (3.2)$$

Suppose further that

$$\text{g.c.d.}\{m \geq 1 : \text{there is an } \epsilon_m > 0 \text{ such that } \sup_{x \in A} P^m(x, \cdot) \geq \epsilon_m \rho(\cdot)\} = 1. \quad (3.3)$$

Then there is a set  $D_0$  such that

$$\pi(D_0) = 1 \text{ and } \sup_{C \in \mathcal{B}} |P^n(x, C) - \pi(C)| \rightarrow 0 \text{ for each } x \in D_0. \quad (3.4)$$

**Theorem 2** Suppose that the Markov chain  $\{X_n\}$  with transition function  $P(x, C)$  satisfies conditions (1.1), (3.1) and (3.2). Then

$$\sup_{C \in \mathcal{B}} \left| \frac{1}{n_0} \sum_{r=0}^{n_0-1} P^{mn_0+r}(x, C) - \pi(C) \right| \rightarrow 0 \text{ as } m \rightarrow \infty \text{ for } [\pi]\text{-almost all } x, \quad (3.5)$$

and hence

$$\sup_{C \in \mathcal{B}} \left| \frac{1}{n} \sum_{j=1}^n P^j(x, C) - \pi(C) \right| \rightarrow 0 \text{ as } n \rightarrow \infty \text{ for } [\pi]\text{-almost all } x. \quad (3.6)$$

Let  $f(x)$  be a measurable function on  $(\mathcal{X}, \mathcal{B})$  such that  $\int \pi(dy) |f(y)| < \infty$ . Then

$$P_x \left\{ \frac{1}{n} \sum_{j=1}^n f(X_j) \rightarrow \int \pi(dy) f(y) \right\} = 1 \text{ for } [\pi]\text{-almost all } x \quad (3.7)$$

and

$$\frac{1}{n} \sum_{j=1}^n E_x(f(X_j)) \rightarrow \int \pi(dy) f(y) = 1 \text{ for } [\pi]\text{-almost all } x. \quad (3.8)$$

Theorem 1 requires condition (3.3), while Theorem 2 does not. Theorem 2 states that if condition (3.3) is violated, one can still apply the Markov chain simulation method, except that one has to work with averages of dependent random variables instead of running a large number of independent chains and working with an (approximately) independent sample. These two theorems are proved in Athreya, Doss, and Sethuraman (1992), where it is also shown that these are the weakest possible conditions that will ensure convergence of a Markov chain for a set of starting points having probability one under the stationary distribution.

There are already many theorems that give conditions that guarantee ergodicity of Markov chains. See the discussion in Section 1 of Athreya, Doss, and Sethuraman (1992). Most of these theorems are stated under two general classes of conditions. Conditions in the first class involve verification of a "recurrence condition" which is much stronger than our condition (3.1). Conditions in the second class involve the stationary distribution of the chain. Since this stationary distribution is unknown, these conditions are difficult to verify. In contrast, our theorems are stated under conditions that involve only the transition function, and thus are, in general, easier to verify.

Theorems 1 and 2 pertain to arbitrary Markov chains. As we mentioned earlier, the Gibbs sampler is the most commonly used Markov chain in Bayesian statistics. We now give a result that facilitates the use of our theorems when the Markov chain used is the Gibbs sampler. We use the notation of Section 1, and assume that for each  $i$ , the conditional distributions  $\pi_{X_i|\{X^{(j)}_{j \neq i}\}}$  have densities, say  $p_{X_i|\{X^{(j)}_{j \neq i}\}}$ , with respect to some dominating measure  $\rho_i$ .

**Theorem 3** Suppose that for each  $i = 1, \dots, k$  there is a set  $A_i$  with  $\rho_i(A_i) > 0$ , and a  $\delta > 0$  such that for each  $i = 1, \dots, k$

$$p_{X_i|\{X^{(j)}_{j \neq i}\}}(x^{(1)}, \dots, x^{(k)}) > 0 \quad (3.9)$$

whenever

$$x^{(1)} \in A_1, \dots, x^{(i)} \in A_i, \text{ and } x^{(i+1)}, \dots, x^{(k)} \text{ are arbitrary,}$$

and

$$p_{X_i|\{X^{(j)}_{j \neq i}\}}(x^{(1)}, \dots, x^{(k)}) > \delta \text{ whenever } x^{(j)} \in A_j, j = 1, \dots, k.$$

Then conditions (3.1) and (3.2) are satisfied with  $n_0 = 1$ . Thus, (3.3) is also satisfied, and the conclusions of Theorems 1 and 2 hold.

We note that condition (3.9) is often checked for all  $x^{(1)}, \dots, x^{(k)}$ .

This theorem is immediate for the case  $k = 2$ . For the general case the proof follows by induction.

## References

- Athreya, K. B., Doss, H., and Sethuraman, J. (1992). A proof of convergence of the Markov chain simulation method. Technical Report No. 868, Department of Statistics, Florida State University.
- Doss, H. and Narasimhan, B. (1993). Bayesian Poisson regression using the Gibbs sampler: A case study in dynamic graphics. Technical Report, Department of Statistics, Florida State University.
- Zaman, A. (1992). Generating random numbers from a unimodal density by cutting corners. Technical Report, Department of Statistics, Florida State University.

# Assessment Of Helicopter Component Statistical Reliability Computations

Donald Neal and William Matthews  
U. S. Army Research Laboratory  
Materials Directorate  
AMSRL-MA-DB  
Arsenal St., Watertown, MA 02172-0001

## Abstract

This report identifies potential errors in computing high statistical reliability for a required component fatigue life. The reliability values were determined from application of a joint probability density (JPD) analysis used in a American Helicopter Society round robin safe life problem.

In the analysis, normal probability density functions(PDFs) were assumed for both the material strength and the spectrum load values. The PDF model parameters were varied and the PDFs were slightly modified (contaminated) in order to examine the sensitivity in computing high statistical reliability when uncertainties exist in assuming the PDF. Lower tails of the PDFs were also modified by truncation; independent of the model contamination, in order to determine the relative influence on reliability from tail modifications as compared with the parameter uncertainties and contamination. The stability of statistical estimates of the extreme tail quantiles and their corresponding probabilities as a function of sample size were examined for a generic distribution.

Assuming a PDF to represent load or material strength is a substantially more critical issue than accurate representation of the extreme lower tail of the PDF when computing high reliability. Sampling trials for extreme tail quantiles and reliabilities indicate that unstable values can result from sample sizes of 100.

The primary conclusion from these analytic results is that the computation of a high statistical reliability may have little or no association with actual engineering high reliability.

## INTRODUCTION

The use of a quantitative high reliability requirement for a helicopter component fatigue design has received considerable attention recently. The U. S. Army established a requirement of .999999 ("six nines") reliability for dynamic components in its most recent helicopter development<sup>1</sup>. Subsequently, the American Helicopter Society Subcommittee on Fatigue and Damage Tolerance, conducted a round robin study of high reliability fatigue methodology applied to a simple structural element<sup>2</sup>. A review of the round robin<sup>3</sup> noted difficulties in the reliability analysis. Each participant used a different fatigue curve and fatigue limit variability which resulted in significantly different fatigue lives, for the six nines reliability requirement. A recent fatigue analysis by a helicopter manufacturer<sup>4</sup> found that "reliability is very sensitive to changes in the population mean strength and scatter". In addition Reference 4 notes "the conclusions of this study are not fully applicable to actual fleet management due to the presence of statistically indeterminant variables such as degraded or non-conforming components".

The present authors in previous study, have investigated the sensitivity of high reliability computations from a stress-strength model<sup>5</sup> to uncertainties in the identification of the probability density functions(PDFs) in the model. The uncertainties are associated with the selection of competing parametric forms( e.g, normal, log-normal, Weibull, etc.) or with the undetected presence of contaminated populations. Contaminated distributions could be bimodal, caused by degraded or non-conforming components, or could be the result of by unexpected loading anomalies. The results from Reference 5 showed that high reliability estimates can vary substantially even for "almost undetectable" differences in the assumed stress and strength PDFs. The authors have also investigated the sensitivity of safe life fatigue reliability of a simple structural element loaded by a simplified spectrum to a variety of uncertainties<sup>6</sup>, demonstrating that a small amount of uncertainty in the parameters of the load or strength PDF resulted in a substantial reduction in the high statistical reliability values for a specified lifetime of the component.

The round robin review, Reference 3, also expressed a concern for the effect of inaccuracies or truncations of the tails of the distributions. An investigation of the truncation of known normal PDF was proposed in order to determine an "acceptable degree of truncation" in computing high statistical reliability. Apparently, this determination would be expected to indicate the portion of the tail region in which an accurate representation of the PDF is not required.

In this report the AHS round robin fatigue problem and its methodology, Reference 2, will be used to investigate the cited issues by considering: a) The effect of small changes in the PDF parameters on the reliability-life relationship. b) The influence upon reliability of the consideration of PDFs which are contaminated, using the methods of Reference 5, by bimodal effects. c) The "true" reliability associated with fatigue lives which have been obtained by satisfying an "apparent" six nines reliability based on normal PDFs which have been truncated in the extreme tail region. d) The relative influence on high statistical reliability of parameter uncertainties, contamination and truncations of the PDF. The consideration of issues involving the extreme tails of the PDFs requires an accurate measure of the truncation point locations, which is difficult to achieve, since sufficient amounts of data is usually not available. In practice, truncation point locations would be estimated from small data sets of load or strength measurements. The stability of the statistical

estimates of the extreme tail quantile and probability values will be investigated based on sampling simulations of a generic normal PDF.

These results will be assessed to indicate the potential role of the PDF tail truncation analysis in providing conclusive information on the acceptability of PDF modeling and whether a quantified .96 reliability provides a meaningful measure which correlates with levels of structural integrity.

## FATIGUE LIFE COMPUTATIONS

The following standardized fatigue life computation procedures were obtained from a round robin study conducted by the AHS, Reference 2. The form of the S-N curve is,

$$N = C(S^* - S_E)^D, \quad (1)$$

where  $N$  = number cycles to failure;  $S_E$  = fatigue strength for very large  $N$  values, for minimum stress equal to zero;  $S^*$  = effective maximum cyclic stress, for minimum stress equal to zero, equivalent to spectrum stresses;  $C$  and  $D$  are parameters from regression least squares analysis.

In order to apply the S-N curve in Equation 1 using the actual operating load spectrum, the following relation for  $S^*$  is required:

$$S^* = \frac{\alpha \cdot S_u \cdot S_L}{S_u - \alpha \cdot S_m + \alpha \cdot S_L/2}. \quad (2)$$

This equation represents a form of the Goodman correction factor used in Reference 2, which converts a defined spectrum mean stress and stress range to an equivalent stress range which causes equal fatigue damage from zero to specified  $S^*$  value.  $S_u$  represents the ultimate strength of the material.  $S_m$  and  $S_L$  represent the mean and range respectively of the nominal stress from a rainflow count obtained in Reference 2, of the standardized Felix 28 spectrum as tabulated in Table 1. The  $\alpha$  value is a scaling parameter for the spectrum load values  $S_L$  and  $S_m$  representing the effective load scaling, over the lifetime of a component. This parameter can provide changes in the baseline spectrum load in order to account for differences in usage, pilot technique, weather, weight, etc.

Let the fatigue life  $N_p$  represent the number of passes prior to the component failure. Then from Miner's Rule, Reference 6,

$$N_p = 1/DF \quad (3)$$

where,

$$DF = \sum_{k=1}^{NK} \frac{n(k)}{N(k)}. \quad (4)$$

The  $n(k)$ s represent the number of cycles for a specific  $k$  value and  $NK$  represents the total number of spectrum load values from Table 1. The  $N(k)$ s are the results from Equation 1.

## RELIABILITY AS A FUNCTION OF LIFE

The following procedures suggested in Reference 1, were applied in order to obtain high reliability( $R$ ) estimates at a specified lifetime. In the analysis the unreliability  $\bar{R}$  is initially determined from application of a discrete joint probability density(JPD) function. The function provides probabilities associated with the simultaneous occurrence of both a spectrum load and a material strength value. In the analysis both the spectrum load scaling factor  $\alpha$  and the material fatigue strength  $s$ , where the mean  $\mu_s = S_E$ , are represented by a normal PDFs as shown in Figure 1. This representation allows for application of the JPD analysis in addition to providing for potential variability in loading and material strength. The scaled version of the spectrum load ( $\alpha S_L$ ) is only involved in computing  $N$  in Equation 1 but not the  $\bar{R}$  estimate. The  $\bar{R}$  computation involving only  $\alpha$  and  $s$  in the JPD computation is therefore simplified. Substitution of  $\alpha$  for  $\alpha S_L$  is valid since both are normally distributed and their probability computations are independent of location. That is, the scaled and unscaled version of the load will provide identical probability estimates with respect to the JPD computation. Since the "event", of identifying a particular value of  $\alpha$  with a component, is independent of the "event" of using a component with particular fatigue strength, the joint probability that  $\alpha = \alpha_i$  and  $s = s_j$  occurs simultaneously can be written as,

$$P(\alpha = \alpha_i, s = s_j) = P(\alpha = \alpha_i) \cdot P(s = s_j), \quad (5)$$

where  $i = 1, 2, 3, \dots, n_1$  and  $j = 1, 2, 3, \dots, n_2$ . The  $n_1$  and  $n_2$  represent the number of events for load and strength respectively. The regions  $A_\alpha$  and  $A_s$  where the events occur which produce higher probability of failure are bounded by the normal PDFs as shown in Figure 1. The load and strength functions are,

$$f_\alpha(\alpha) = \frac{1}{\sigma_\alpha \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{\alpha - \mu_\alpha}{\sigma_\alpha} \right)^2}, \quad (6)$$

and

$$f_s(s) = \frac{1}{\sigma_s \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{s - \mu_s}{\sigma_s} \right)^2}, \quad (7)$$

where  $(\mu_\alpha, \sigma_\alpha)$  and  $(\mu_s, \sigma_s)$  are the population means and standard deviations for the load and strength, respectively. Referring to Figure 1 and Equation 5, the JPD can be written as

$$P_{i,j} = P_{\alpha_i} \cdot P_{s_j}, \quad (8)$$

where,

$$P_{\alpha_i} = \Delta \alpha_i \cdot f_\alpha(\bar{\alpha}_i), \quad (9)$$

$$P_{s_j} = \Delta s_j \cdot f_s(\bar{s}_j), \quad (10)$$

and  $i = 1, 2, 3, \dots, n_1$  and  $j = 1, 2, 3, \dots, n_2$ . After determining the joint probability values  $P_{i,j}$  from Equation 8, the corresponding  $\alpha$  and  $s$  associated with these probabilities are introduced in Equations 2 and 1 respectively. This determines a specific number ( $N_{i,j}$ ) cycles to failure of the material for the corresponding probability ( $P_{i,j}$ ) of the joint occurrence of  $\alpha_i$  and  $s_j$ . The lifetime estimate  $N_P$  from Equation 3 for the joint  $\alpha_i$  and  $s_j$  event is obtained from the following application of the spectrum load data  $\{S_L(k)\}_1^{NK}$ ,  $\{S_m(k)\}_1^{NK}$  and  $\{n(k)\}_1^{NK}$  in Table 1, where  $NK$  is the number

of spectrum loads. The damage fraction for a specified event can be determined from Equation 4 and written as

$$DF_{ij} = \sum_{k=1}^{NK} \frac{n(k)}{N_{ij}(k)} \quad (11)$$

where  $n(k)$ s are the spectrum load cycles corresponding to the original tabulated loads  $S_L(k)$ . From Equations 3 and 8, the lifetime values are then computed from

$$N_P(ij) = 1/DF_{ij} \quad (12)$$

These values correspond to the joint probability that  $\alpha = \alpha_i$  and  $s = s_j$ . The above process is repeated  $M = n_1 n_2$  times, where  $n_1$  and  $n_2$  represent the number of mesh points associated with the tail region of the PDFs in Figure 1. All combinations of  $i$  and  $j$  are introduced in order to obtain paired  $P_{ij}$  and  $N_P(ij)$  values. Ordering only the  $N_P$  values from the smallest to largest and retaining their original corresponding  $P_{ij}$  probabilities describes a discrete PDF representing the component probability of failure  $P_f(t)$  as an array of lifetimes  $\{N_P^*(t)\}_1^M$ , where  $t$  is an integer defining the ordering of the  $N_P^*$  values. See a graphical display of a PDF in Figure 2. In order to obtain the unreliability  $\bar{R}$  for a given  $t_1$  in Figure 2, a cumulative density computation is required. This is accomplished by selecting an ordered value from  $N_P^*$  and computing the sum

$$\bar{R}(t_1) = \sum_{t=1}^{t_1} P_f(t) \quad (13)$$

Note, the reliability  $R$  can be obtained from  $R = 1 - \bar{R}$  and the lifetime values can be determined from a given  $\bar{R}$ .

## CONTAMINATED PROBABILITY DENSITY FUNCTIONS

In order to illustrate the sensitivity of high reliability calculations to small deviations from the assumed models, the approach taken in Reference 5, is applied. Consider the situation where with a high probability of  $1 - \epsilon$ , samples are obtained from a primary PDF, while with probability  $\epsilon$  samples come from a secondary PDF. This bimodal probability model is a type of a general class referred to as a *contaminated* models. The secondary component is called the *contamination* and the probability  $\epsilon$  is the *amount* of contamination. An example may help clarify this idea. Consider the situation where 99% of the specimens are obtained from a population of "good" specimens while the remaining 1% of the time consistently lower strength measurements are obtained, either due to manufacturing defects or to faulty testing. The primary PDF would correspond to the "good" specimens, the contamination would represent the distribution of flawed specimens, and the amount of contamination is  $\epsilon = 0.01$ . The following procedure is introduced in order to examine the effects of computing high reliability values when uncertainties exist in selecting the PDFs for the joint density computations. Initially, values are obtained from the JPD computation using PDFs  $f_\alpha$  and  $f_s$  in Equations 6 and 7. Another  $R$  value is then obtained by applying the PDFs with a small amount of contamination  $\epsilon$ .

The  $f_s$  PDF with variance contamination for the strength data is written as,

$$f_{cs}^v = (1 - \epsilon)f_s(\mu_s, \sigma_s^2) + \epsilon f_s(\mu_s, k_1^2 \sigma_s^2), \quad (14)$$



where  $\mu_s$  and  $\sigma_s$  are defined in Equation 7,  $k_1$  is a scaling factor and  $100\epsilon$  is the percent contamination. A similar contaminated distribution for  $f_\alpha$  representing load  $\alpha$  can be written as

$$f_{c\alpha}^v = (1 - \epsilon)f_\alpha(\mu_\alpha, \sigma_\alpha^2) + \epsilon f_\alpha(\mu_\alpha, k_1^2 \sigma_\alpha^2), \quad (15)$$

with scaling factor  $k_1$  and  $\mu_\alpha$  and  $\sigma_\alpha$  obtained from Equation 6. Variance contamination produces effects which can be considered to represent uncertainties associated with the selection of competing PDF models.

A strength distribution with mean location contamination is

$$f_{cs}^L = (1 - \epsilon)f_s(\mu_s, \sigma_s^2) + \epsilon f_s(\mu_s \pm k_2 \sigma_s, \sigma_s^{*2}), \quad (16)$$

where  $k_2$  is a scaling factor for  $\mu_s$  and the sign determines which tail of the function is to be contaminated and  $\sigma_s^{*2}$  is the variance for  $\mu_s \pm k_2 \sigma_s$ . The contaminated function for load  $\alpha$  can be determine in a similar manner. The location contaminated PDF can represent the rare occurrence of exceptionally high loads or the unusually low material strength of a degraded or non-conforming component in computing the reliability. For  $\epsilon = .01$  and  $k_1 = 4$ , graphical results in Figure 3, show an almost undetectable difference between the original normal PDF and the contaminated one. A linear relationship to obtain R from the JPD function application can be obtained by combining both contaminated and uncontaminated functions such that,

$$R^* = (1 - \epsilon_1)(1 - \epsilon_2)R_{00} + \epsilon_1(1 - \epsilon_2)R_{10} + \epsilon_2(1 - \epsilon_1)R_{01} + \epsilon_2\epsilon_1R_{11}, \quad (17)$$

where  $100\epsilon_1$  and  $100\epsilon_2$  are the percent contamination in the  $\alpha$  and  $s$  distributions and  $R_{mn}$  represents reliabilities obtained from contaminated conditions designated by  $m$  and  $n$ . If  $m, n = 0$ , then there is no contamination. If  $m = 1$  and  $n = 0$  then  $f_\alpha$  is contaminated. If  $m, n = 1$ , then both  $f_\alpha$  and  $f_s$  are contaminated. For example, if there is contamination of the strength PDF with respect to the variance then,

$$R^* = (1 - \epsilon_2)R_{00} + \epsilon_2R_{01}. \quad (18)$$

The  $R_{mn}$  values are obtained from  $\bar{R}$  in Equation 13. This procedure provides an effective approach for demonstrating the effects of PDF uncertainties in determining high R values.

## MODIFYING TAILS OF THE PDFS

A modification of the  $f_\alpha$  and  $f_s$  PDFs' upper and lower tail regions respectively was introduced in the analysis to investigate truncation effects as suggested in Reference 3. A proposed modification<sup>7</sup> is shown in Figure 4. The lumping method of truncation shown in the figure was selected so that the area under the modified PDF remains equal to one. This was accomplished by determining the area under extreme tail regions associated with the probabilities  $P_{\alpha_{n_1}}$  and  $P_{s_{n_2}}$  obtained from Equations 20 and 21. These areas were lumped at the truncation points  $z_1$  and  $z_2$  for  $\alpha$  and  $s$  and the reliability  $R_L$  values were determined from Equation 13, with the lumped  $f_\alpha$ ,  $f_s$ . A comparison was then made between lumped and unlumped results in order to determine if the effects of the uncertainties in the extreme tails are significant in computing high R values. The modification also represents a substantial difference in the lower tail region when compared with the original PDFs.

The R computation using the lumped PDFs, in Figure 4, is the same as that described previously, except in Equation 8,

$$P_{n_1 n_2} = P_{\alpha_{n_1}} \cdot P_{s_{n_2}}, \quad (19)$$

where,

$$P_{\alpha_{n_1}} = \int_{z_1}^{\infty} f_{\alpha} \cdot d\alpha, \quad (20)$$

$$P_{s_{n_2}} = \int_{-\infty}^{z_2} f_s \cdot ds, \quad (21)$$

and  $z_1 = \mu_{\alpha} + \lambda_1 \sigma_{\alpha}$ ,  $z_2 = \mu_s + \lambda_2 \sigma_s$ , are the truncation points of the PDFs shown in Figure 4. The  $f_{\alpha}$  and  $f_s$  PDFs are defined in Equations 6 and 7. The R computation procedures are then applied using the newly defined  $P_{\alpha_{n_1}}$  and  $P_{s_{n_2}}$  values.

## PDF PROBABILITY AT TRUNCATION POINTS

The lumping procedure described previously was introduced in order to determine the effects on the R values from modifications to the extreme tail of the PDFs.

In applying Equations 20 and 21, it is assumed that the PDFs and the truncation points  $z_1$  and  $z_2$  are known exactly; which is usually not the case for engineering problems involving material strength or loading measurements. Since the accuracy in estimating the truncation point locations is essential in determining the importance of correct extreme tail representation in obtaining high R values, the following study was performed involving determination of the reliability and quantile values at selected truncation points as a function of sample size. Quantile values of  $f_s$  can be written as,

$$S_q = \mu_s - K \cdot \sigma_s, \quad (22)$$

where  $\mu_s$  and  $\sigma_s$  are known mean and standard deviation of  $f_s$ . The K values can be selected to define points where truncation may be introduced in the high reliability computation. Unfortunately, the  $\mu_s$  and  $\sigma_s$  values are not known sufficiently well for an accurate measurement of  $S_q$  unless very large data sets are applied.

The following simulation process examines the sample size ( $n_2$ ) effects in computing the truncation points. In the simulation process, a  $n_2$  set of  $s_i$  normally distributed values are selected from

$$s_i(i) = \mu_s(1 + v_s \cdot Q_i), i = 1, 2, 3, \dots, n_2 \quad (23)$$

where the  $Q_i$  values are obtained from a standard normal distribution with  $\mu_s = 100$  and  $v_s = .10$ . The  $v_s$  value is the CV =  $\sigma_s/\mu_s$  and  $\mu_s$  is the population mean. From the  $s_i$  values the mean  $\bar{s}_i$  and variance  $VAR(s_i)$  are determined. An estimate of the population quantile  $S_q$  is then

$$\hat{S}_q = \bar{s}_i - K \cdot (VAR(s_i))^{1/2} \quad (24)$$

The probability  $P_T$ , where  $P_T$  equals  $\text{Prob}[s_i \geq S_q]$  can be estimated by the proportion of  $s_i$  values which are greater than  $S_q$ . The process involving Equations 23 and 24 is repeated many times so that

the  $P_T$  and  $S_q$  values are not effected by further increasing the number of simulations. The range of  $P_T$  and  $S_q$  values is a measure of the statistical stability of the sampling process. Particular quantile values such as 1% or 99% can be obtained by forming a cumulative probability distribution for  $S_q$  and  $P_T$ . The probabilities are ordered from the smallest to largest and their percent is determined from their numbered position in the ordering which is divided by the total number of simulations. The  $S_q$  quantile is the value at the same numbered position of interest.

## RESULTS AND DISCUSSION

In this section, results are obtained for the AHS round robin problem in Reference 2. A thin AISI 4340 steel plate with a central hole is loaded by the Felix 28 spectrum. The ASTD S - N curve coefficients are used:  $C = 3.5 \times 10^6$ ,  $D = -1.47164$  and  $S_E = 54.5$  KSI. The  $S_U$  value is 180 KSI in Equation 2. The number of mesh points  $n_1$  and  $n_2$  of the PDFs in the JDF computation (Equation 5) are each 50, where  $C_\alpha = \mu_\alpha$  and  $C_s = \mu_s$  (see Figure 1).

In Figure 5, representation of reliability as a function of life ( $N_P$ ) is shown for selected CV values used in defining the PDFs  $f_\alpha$  and  $f_s$  in the R computation, with a mean load factor ( $\mu_\alpha$ ) of .70. The results show a very rapid initial decrease in reliability followed by a more moderate decline in reliability as the lifetime values increase. Results were the same for other  $\mu_\alpha$  values. R estimates also decreased with an increase in the assumed CV value. For example, a CV = .05 for both PDFs provides a much greater  $R(.9_{(11)})$  estimate than for a CV = .07 which is  $.9_{(6)}$  when  $N_P = 100$ . The results designated by the \* were obtained from applying lumped PDFs for  $\lambda_1 = \lambda_2 = 3.5$ . These values of  $\lambda$  are almost the maximum amount of truncations that will provide the  $R = .9_6$  requirement. The maximum truncation was avoided because the  $R = .9_6$  would be met at the discrete probability value of the discrete joint distribution which includes the lumped values. The  $R = .9_6$  value from the lumped PDF differs slightly from the unlumped PDF where  $R = .9_{54}$  when  $N_P = 50$  and the CV = .8. This result indicates that modifications of the extreme lower tails of the PDFs do not cause large differences in computing high reliability.

Figure 6 shows reliability as a function of life for selected  $\mu_\alpha$  load factors. As in Figure 5, there is a reduction in reliability with an increase in  $N_P$  values. There is also an obvious decrease in reliability with an increase in load factor. In the case where  $N_P = 275$ , an  $\mu_\alpha$  of .5 and .6 resulted in  $R = .9_{11}$  and  $.9_6$  respectively. This shows a substantial sensitivity in computing R for a relatively small differences which may occur in estimating  $\mu_\alpha$ .

Figure 7 also shows the reliability values as a function of life ( $N_P$ ). The assumed CV value for the PDFs is .07 and applied loads are  $\mu_\alpha = .5$  and .7. The dash lines represent results from applying the contaminated PDFs described in Equations 14 and 15, for the case of 1% contamination in both PDFs and  $k_1 = 4$ . This almost undetectable level of contamination caused a drastic reduction in reliability for  $R > .9_6$ . The results from the contaminated PDFs application which represent the uncertainties in assuming a specific function, demonstrate the importance of identifying precise PDFs in computing component high statistical reliability. In contrast, at  $N_P = 130$ , the unlumped result for of  $.9_{52}$  is only moderately reduced from the value of  $.9_6$  obtained for lumped PDFs for  $\lambda_1$

and  $\lambda_2 = 3.5$ .

In Table 2, reliability results are tabulated using both lumped( $R_L$ ) and unlumped( $R_{UL}$ ) PDF applications for selected CV values. These results show a reduction in  $R_L$  and  $R_{UL}$  values with increasing CV values. In the case of CV = .05 and .06, the  $R_L$  estimate of .976 was the maximum obtainable because of the discrete nature of the PDF truncation procedures. Comparing the values of  $R_L = .976$  with  $R_{UL} = .96$  for CV = .07 shows a relatively small difference in the reliability values. This result indicates that uncertainties in modeling the extreme lower and upper tails of the PDFs cause relatively small differences in computing high R values. The issue that is important involves the substantial reduction in the  $R_L$  and  $R_{UL}$  values with increasing CVs. Since the CV values are often estimated from coupon data that are assumed to be relevant to actual component behavior, substantial uncertainties can result in estimating R.

Table 3 shows effects in computing R from applying both contaminated and uncontaminated PDF applications. In the case where  $\mu_\alpha = .5$  and  $N_P = 80$ , the uncontaminated PDF result is  $R \gg .912$ . When there is a 1% contamination of PDF for the load, the value is reduced to .96. Contaminating the strength PDF by 1% resulted in a reduction from twelve nines(uncontaminated) to three nines(contaminated) in the R estimate. Very similar results were obtained for contamination of both PDFs. As previously shown in Figure 7, obtaining extremely high reliability greater than twelve nines will not provide the necessary conservative estimate for R if there is even a very small amount of uncertainty in assuming a PDF in the R computation. The case where  $\mu_\alpha = .7$  represents application of both the lumped and contaminated PDFs in the R computation. The lumped PDF result without contamination showed  $R = .976$  which is reduced to .978955 when the contaminated PDF was also applied. Again, this substantial reduction in R demonstrates the importance in the accuracy of the PDF. Potential uncertainties associated with defining the extreme tails of the PDFs become insignificant relative to the accuracy of PDF assumption in computing high R values. The table also shows that by increasing the CV value the  $N_P$  value is reduced but the reduction in R from the contaminated PDF are the same as those for CV = .07. Summarizing the results in the table: it is critical in computing high statistical R values that the PDFs are known almost exactly while uncertainties in the extreme tails of the PDFs are relatively insignificant.

Table 4 provides results similar to those in Table 2 except that the load  $\mu_\alpha$  is varied in order to examine the effects of uncertainty in determining the mean scaling load factor in the reliability - life estimating process. The results again show a substantial difference in R for an uncertainty in  $\mu_\alpha$ . For example, when  $\mu_\alpha = .7$ ,  $R = .96$  and for  $\mu_\alpha = .8$ ,  $R = .93662$  which implies that there will be one failure and 338 failures in a million for loads  $\mu_\alpha = .7$  and .8 respectively. This substantial difference relative to uncertainties in estimating  $\mu_\alpha$  indicates that the R computations are very sensitive to uncertainties in the load. The table shows little difference between  $R_{UL}$  and  $R_L$  for  $\mu_\alpha \geq .7$ . When  $\mu_\alpha < .7$  no quantitative comparisons are possible because of the PDF truncations. The R value for  $\mu_\alpha = .7$  shows a substantial decrease in  $R_{UL}$  from approximately .96 to .9789 for the uncontaminated and contaminated PDFs respectively. A similar result is shown for  $R_L$  at  $\mu_\alpha = .7$ . The  $R_{UL}$  and  $R_L$  results at  $\mu_\alpha = .7$  showed a small effect on R with a substantial modification of the extreme tails of PDFs. This shows that the PDF assumption is critical in computing R while accuracy in representing the extreme tail of the PDFs is much less critical.

In Table 5 are the results of sampling a generic normal PDF to examine the stability of the statistical estimates at the potential truncation points. The median probability and quantile values are shown for a range of sample sizes ( $n_2$ ). Included in the results are the upper and lower bounds on the 98% confidence interval on the median estimates. Reliability ( $P_T$ ) values are obtained at  $K = 3.5$  and  $4.75$  where truncation may be introduced in high reliability computations in studying effects in PDF tail modifications. This was done in order to examine if there is instability in  $P_T$  at the points due to the sample size. The sampling trials were repeated 6000 times which was sufficient to ensure that the tabulated values would not change with additional trials. Results from the table indicate that relatively unstable  $P_T$  values will be obtained for even a sample size of 100. In this case, the true  $P_T$  value is  $.93767$  for  $K = 3.5$  but the simulation result shows an inner confidence range of  $.928172$  to  $.9482$  for  $P_T$  values associated with  $S_q$  estimates. This uncertainty in the  $S_q$  location, reduces the validity in assuming that if lumping a PDF does not cause a substantial change in  $R$ , then the PDF will be adequate for computing high  $R$  values such as  $.96$ . Another more obvious example is the case where  $n_2 = 6$  which shows an interval of  $.835123$  to  $.998$  for  $K = 3.5$  and  $.918483$  to  $1.0$  when  $K = 4.75$ . The inner 98% ranges of  $S_q$  quantile values for  $K = 4.75$  and  $K = 3.5$  also show a substantial overlap. This case shows that even if the lumping process provides results showing small differences in  $R$  between  $K = 3.5$  and  $4.75$  (using an unverified assumed known truncation point), the inference is meaningless. That is, the substantial uncertainty associated with computing  $R$  at unverified truncation points prevents making any assessment regarding the need for accurate representation of the extreme tail of the PDF in computing  $.96$   $R$  values.

These results are consistent with results of truncations of normal PDFs<sup>8</sup> where, for truncations of less than 10 to 20 percent of the population, quantiles would fall within permissible limits of random variation, unless sample sizes are very large. Reference 5 shows various levels of uncertainty associated with computing high reliability from a stress-strength statistical model as a function of sample size. These results relate directly to the sample size issue discussed in this report.

The substantial sensitivities of  $R$  in each of the figures and tables relate to uncertainties in only one parameter, while the others are held constant. In design, the uncertainties in more than one parameter such as  $\mu_\sigma$  and  $CV$  could cause increased  $R$  sensitivity. There are many complex issues involved in obtaining a component population PDF for effective load severity scaling parameter, over lifetime. There is no industry standard approach to characterizing the load history and limited experience in determining loading PDFs. Therefore, the substantial influence on high reliability caused by loading PDF uncertainties could cause a serious problem in the implementation of a high reliability requirement.

These results are based on a single S-N curve, in contrast to the AHS round robin problem in which each participant used a different S-N curve. Thus, very substantial variability in the  $R$ -lifetime relationship can be expected even when the S-N curve shape and mean fatigue limit stress is fixed.

These results and the previous analyses of contaminated PDFs in Reference 5, support the concern expressed in Reference 4, for the issue of a decrease in reliability caused by degraded or non-conforming components. The approach of attempting to obtain statistically very high  $R$  values ( $.912$ ) to compensate for uncertainties in assuming a PDF may not provide an effective margin of safety

or conservatism relative to a .96 requirement.

The comparisons between  $R_L$  and  $R_{UL}$  do not directly relate to PDF modeling for design. In the approach used in this report the lumped value in the PDF is made exactly equal to the extreme tail of a known PDF with which it is being compared. In the design process, the difference of interest is between a truncated assumed PDF and a "correct" PDF which is unknown. The lumping approach used in this study would tend to minimize the difference between  $R_L$  and  $R_{UL}$  relative to an actual design process.

No conclusion can be reached about an acceptable degree of truncation from this study. For  $\lambda_1$  and  $\lambda_2 \leq 3.5$  it appears that truncation is not acceptable for the .96 requirement. Variation in  $R$  from less than one "nine" to values approaching two "nines" were obtained for idealized conditions which minimize reliability differences as noted previously. For  $\lambda_1$  and  $\lambda_2 > 3.5$ , the sampling results indicate that it does not appear to be feasible to obtain satisfactory representation of PDF unless very large data sets are available. More important, the issue of acceptable degree of truncation appears to be of secondary importance relative to the sensitivity of high  $R$  to the expected uncertainties in assuming a specific PDF representations.

## CONCLUSIONS

Unstable high statistical reliability values for a fatigue loaded component can result from uncertainties in assuming the PDF model and determining its parameters without using very large data sets in the analysis.

Estimates of the extreme tail quantiles and their corresponding reliabilities can be unstable unless large data sets are used.

Analysis of the effects of extreme tail modification does not provide decisive information on the adequacy of PDF modeling. Tail modification effects on reliability are small relative to the effects of uncertainties in assuming a PDF model.

*The primary conclusion, from the analytic evaluation in this report, is that computation of high statistical reliability may have little or no association with actual component reliability.*

## REFERENCES

1. Arden, R.W. and Immen, F.H., *U.S Army Requirements For Fatigue Integrity*, Proceedings of American Helicopter Society National Technical Specialists Meeting On Advanced Rotorcraft Structures, Williamsburg, VA, October 1990.
2. Everett, R.A., Bartlett, F.D., and Elber, W., *Probabilistic Fatigue Methodology For Six Nines Reliability*, AVSCOM Technical Report 90-B-009, NASA Technical Memorandum 102757, December 1990.
3. Schneider, G. and Gunsallus, C., *Continuation Of The AHS Round Robin On Fatigue and Damage Tolerance*, Presented At The American Helicopter Society Forty Seventh Annual Forum, Phoenix, AZ, May 1991.
4. Thompson, A.E and Adams, D.O. *A Computational Method For The Determination Of Structural Reliability Of Helicopter Dynamic Components*, Presented At The American Helicopter Society Annual Forum, Washington, D.C., May 1990.
5. Neal, D.M., Matthews, W.T. and Vangel, M.G., *Model Sensitivity In Stress-Strength Reliability Computations*, U.S. Army Materials Technology Laboratory, MTL TR 91-3, January 1991.
6. Neal, D.M., Matthews, W.T., Vangel, M.G. and Rudalevige, T., *A Sensitivity Analysis On Component Reliability From Fatigue Life Computations*, U.S. Army Materials Technology Laboratory MTL TR 92-5, February 1992.
7. Gunsallus, C., Memorandum To American Helicopter Society Fatigue And Damage Tolerance Subcommittee, August 28, 1991.
8. Hald, A., *Statistical Theory With Engineering Applications* Wiley, 1952, P. 146.

Table 1: Rainflow low-high load sequence derived from Felix-28

$k$	$S_L$	$S_m$	$n(k)$	$k$	$S_L$	$S_m$	$n(k)$
1	2.80	25.59	354	26	42.63	29.21	207
2	2.80	32.83	334	27	42.63	36.45	1274
3	6.42	29.21	416	28	46.25	21.97	274
4	10.04	29.21	609	29	46.25	25.59	6239
5	10.04	36.45	1228	30	46.25	29.21	4274
6	10.04	40.07	810	31	46.25	40.07	604
7	13.66	36.45	2	32	49.87	3.86	268
8	17.28	18.35	140	33	49.87	25.59	956
9	17.28	32.83	78	34	49.87	29.21	2179
10	20.91	32.83	2061	35	53.49	25.59	2
11	20.91	36.45	90	36	53.49	29.21	116
12	24.53	-7.00	140	37	57.12	25.59	5
13	24.53	18.35	140	38	57.12	29.21	185
14	24.53	36.45	2040	39	60.74	29.21	25
15	28.15	29.21	833	40	64.36	25.59	7
16	31.77	25.59	346	41	64.36	29.21	8
17	35.39	25.59	7904	42	64.36	32.83	75
18	35.39	29.21	56	43	67.98	29.21	9
19	35.39	32.83	71072	44	71.60	29.21	16
20	39.39	43.69	2529	45	75.22	25.59	7
21	39.01	21.97	3014	46	78.84	18.35	5
22	39.01	25.59	42825	47	78.84	25.59	1
23	39.01	29.21	6393	48	82.46	21.97	128
24	39.01	43.69	252	49	82.46	29.21	16
25	42.63	25.59	480	50	89.70	25.59	8



Table 2: Reliability from lumped and un-lumped PDFs as a function of the CVs

CVs	$R_L$	$R_{UL}$
0.05	> .976	.911
0.06	> .976	.97
0.07	.976	.96
0.08	.954	.9483
0.09	.9430	.93885
0.10	.93646	.93533

$R_L$  - Reliability from lumping PDFs

$R_{UL}$  - Reliability from normal PDFs

CVs - Coefficient of variations

$N_P = 80$  Passes,  $\mu_\alpha = .70$  Load factor,  $S_E = 54.5$  KSI Strength

Table 3: Effects of individual PDF contaminations on reliability estimates

$\mu_\alpha$	CV	$N_P$	$R_U$	$R_{CL}$	$R_{CS}$	$R_{CLS}$
.5	.07	80.0	> .912	.96	.93816	.93814
.7	.07	80.0	> .96	.93730	.93186	.928914
.7*	.07	80.0	> .976	.93733	.93224	.928955
.7	.10	4.0	> .96	.93848	.93092	.928940

\* Both tails of PDFs are lumped at  $3.5 \sigma$

$R_U$  - Reliability uncontaminated PDFs

$R_{CL}$  - Reliability contaminated(1%) load( $\alpha$ ) PDF

$R_{CS}$  - Reliability contaminated(1%) strength( $S_E$ ) PDF

$R_{CLS}$  - Reliability from both load and strength PDFs contamination

Table 4: Reliability from lumped and unlumped PDFs VS. Load  $\mu_\alpha$

$\mu_{\alpha}$	$R_{UL}$	$R_L$
0.4	$> .9_{12}$	$> .9_7$
0.5	$> .9_{12} [.999814]^*$	$.9_7 130$
0.6	$.9_{10}$	$.9_7 297$
0.7	$.9_6 [.998914]^*$	$.9_7 [.998955]^*$
0.8	$.999662$	$.999724$
0.9	$.984984$	$.984710$
1.0	$.855658$	$.851946$

$R_L$  - Reliability from lumping PDFs

$R_{UL}$  - Reliability from unlumped PDFs

\* Results from contaminated PDFs

$N_P = 80$  Passes And  $CV_s = .07$

Table 5: Confidence interval on  $P_T$  and  $S_q$  quantile

$N$	Normal PDF $\mu = 100, CV = .10$					
	$K = 3.5$			$K = 4.75$		
	Median	Lower Bound	Upper Bound	Median	Lower Bound	Upper Bound
6	$.999480^*$ (67.07)**	$.835123$ (90.82)	$.9_8 78$ (37.23)	$.9_5 58$ (55.10)	$.918483$ (86.67)	$1.00$ (16.56)
10	$.999669$ (66.00)	$.925734$ (85.04)	$.9_7 86$ (43.75)	$.9_5 81$ (54.04)	$.987086$ (78.07)	$.9_{13}$ (24.78)
20	$.999721$ (65.48)	$.981327$ (78.92)	$.9_6 6$ (50.64)	$.9_5 84$ (53.36)	$.998403$ (70.56)	$.9_{10} 7$ (34.27)
50	$.999745$ (65.36)	$.995914$ (73.92)	$.9_5 4$ (56.13)	$.9_5 9$ (52.81)	$.999845$ (63.95)	$.9_8 81$ (41.14)
100	$.999760$ (65.13)	$.998172$ (71.05)	$.9_4 82$ (59.01)	$.9_5 9$ (52.68)	$.9_4 60$ (60.52)	$.9_8$ (43.74)
1000	$.999765$ (65.02)	$.999524$ (66.96)	$.999890$ (63.07)	$.9_6$ (52.52)	$.9_5 6$ (55.11)	$.9_6 7$ (49.99)

\*  $P_T$  Probability

\*\* Corresponding  $S_q$  quantile value

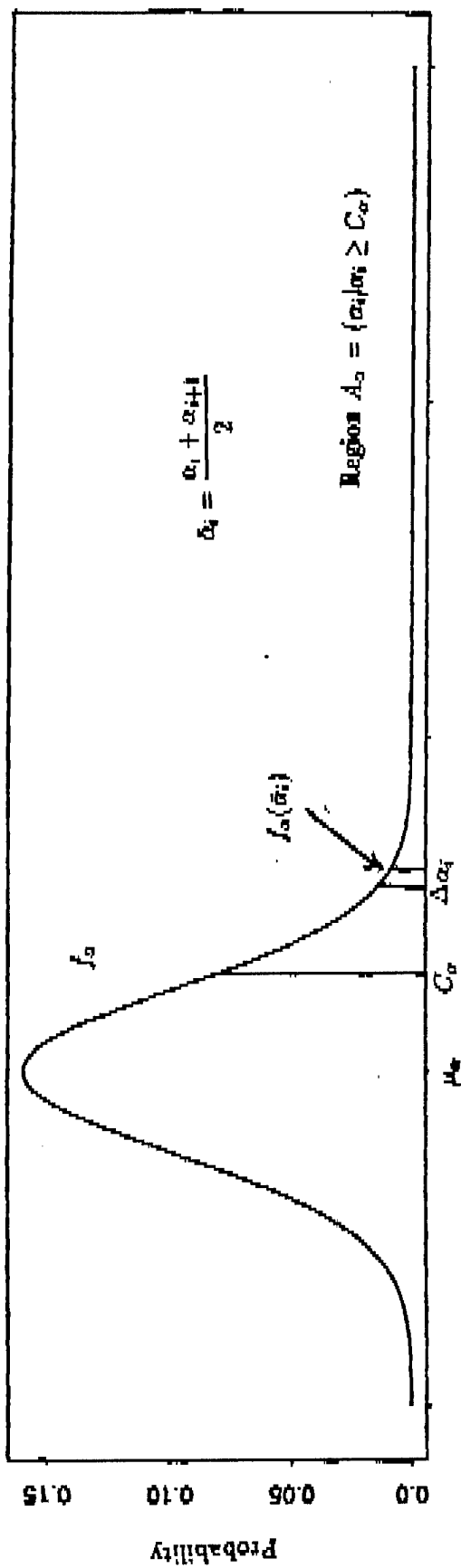


Figure 1a. Normal PDF for a load - JPF Region  $A_2$

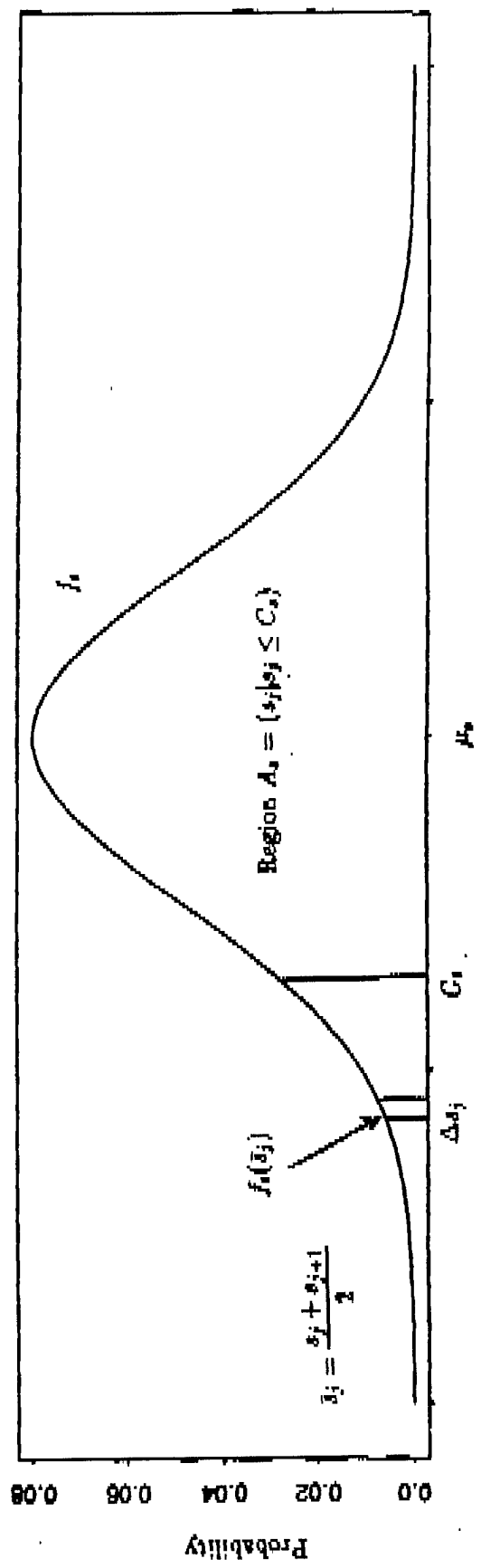


Figure 1b. Normal PDF for strength  $s$  and JPF Region  $A_1$

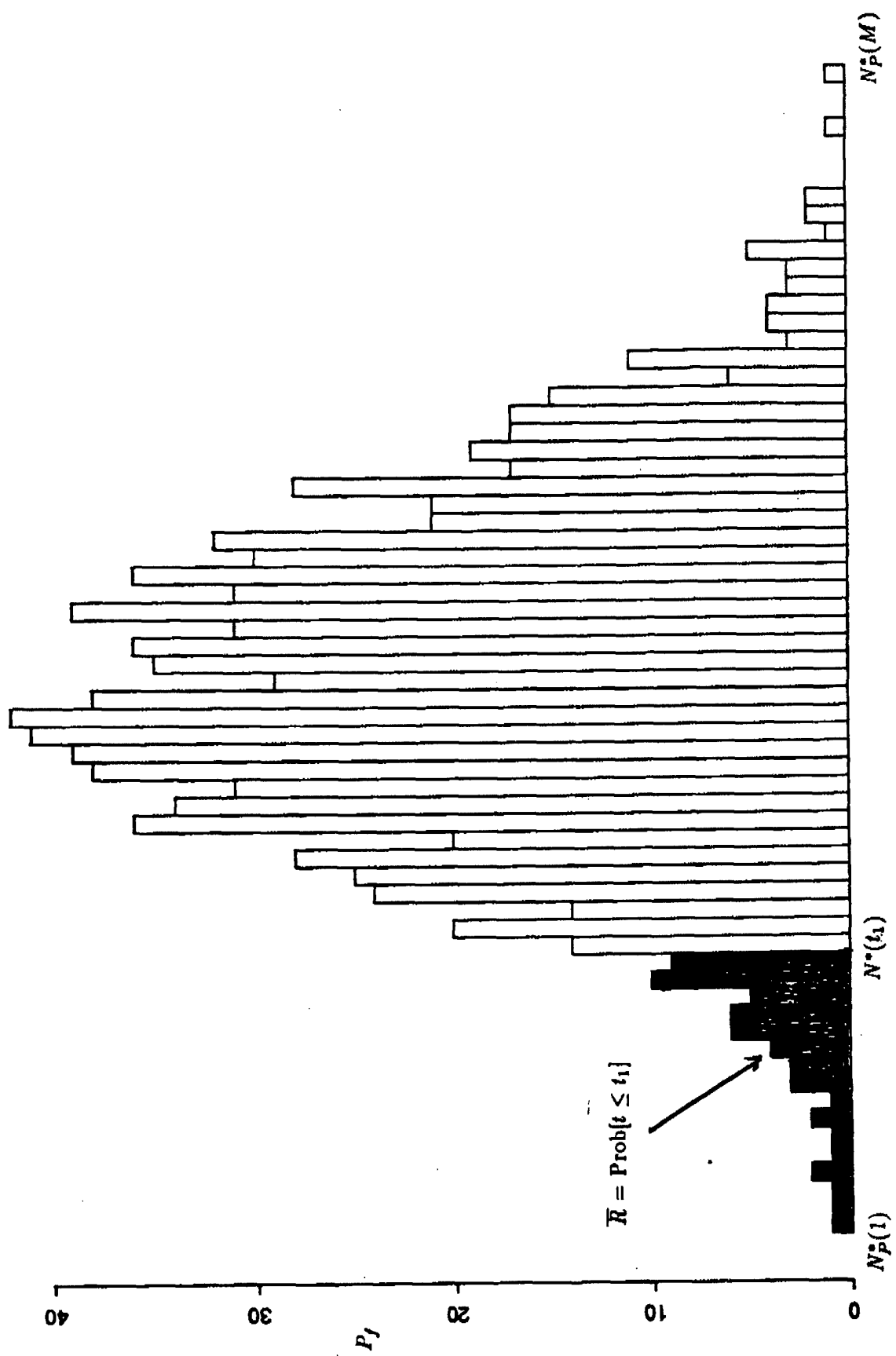


Figure 2. Discrete PDF for unreliability( $\bar{R}$ ) values.

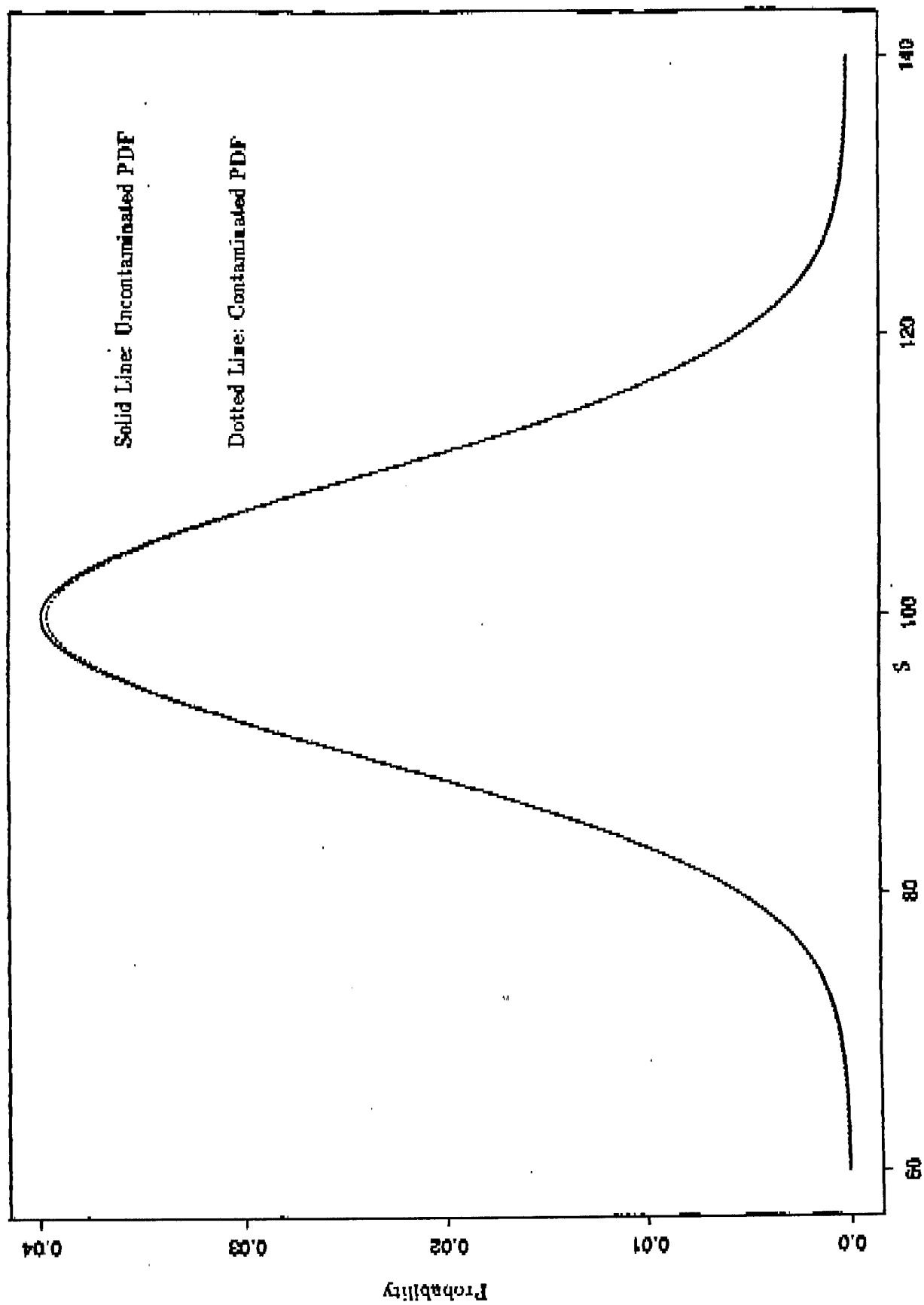
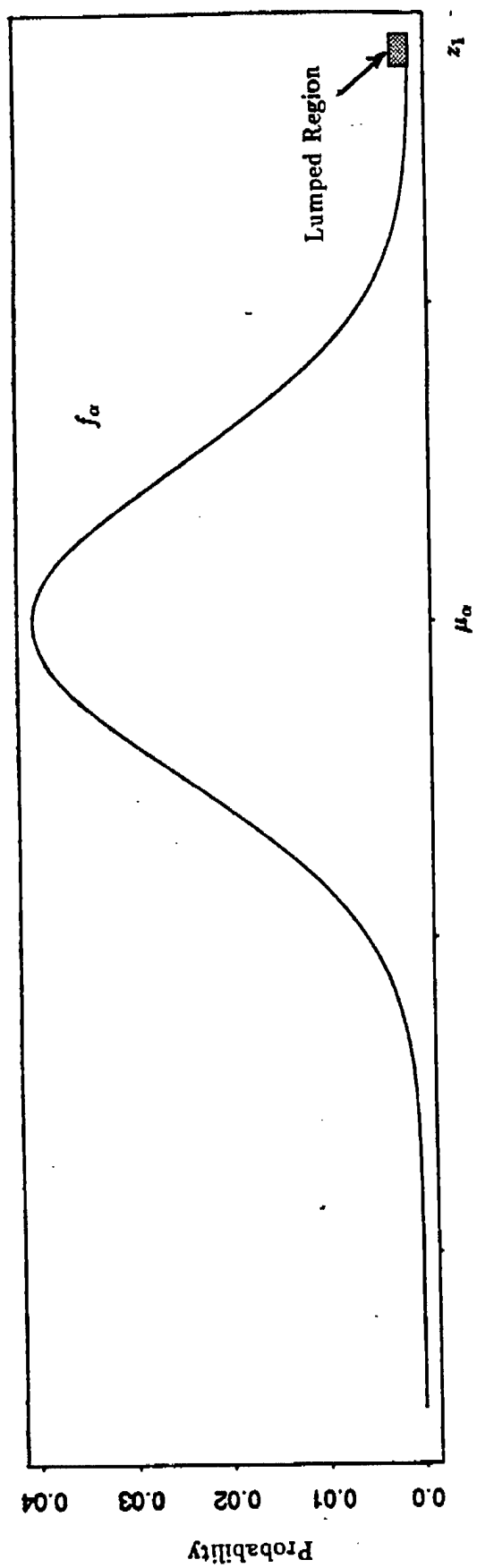


Figure 3. Contaminated(1%) and uncontaminated normal PDFs



682

Figure 4a. Lumped normal load( $\alpha$ ) PDF

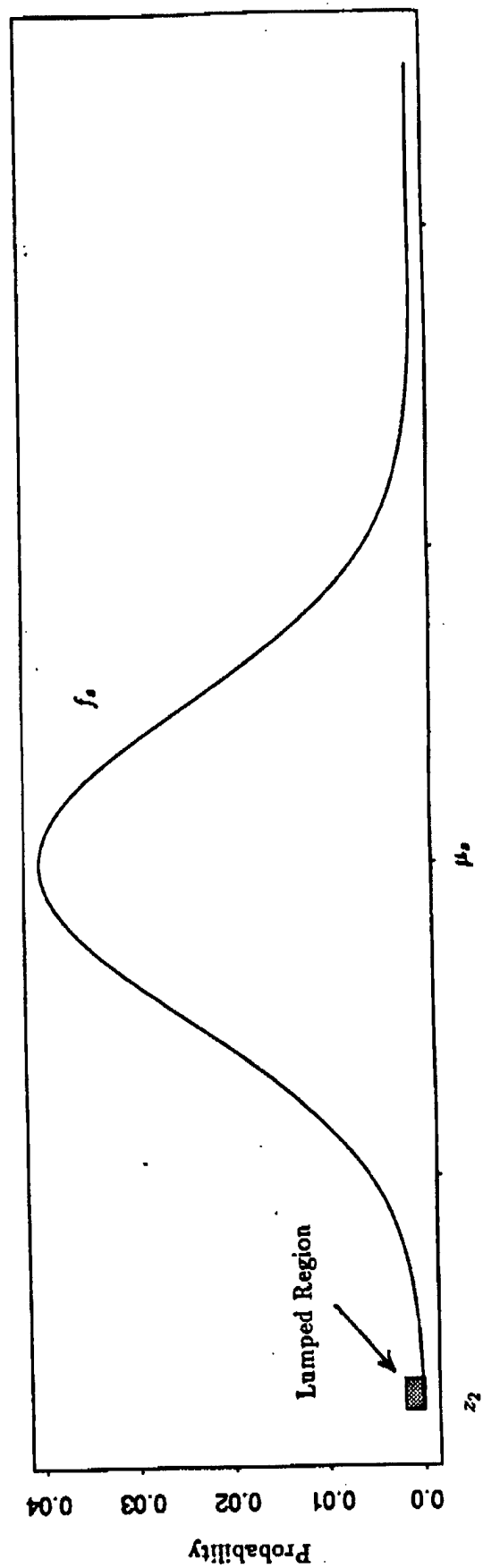


Figure 4b. Lumped normal strength( $s$ ) PDF

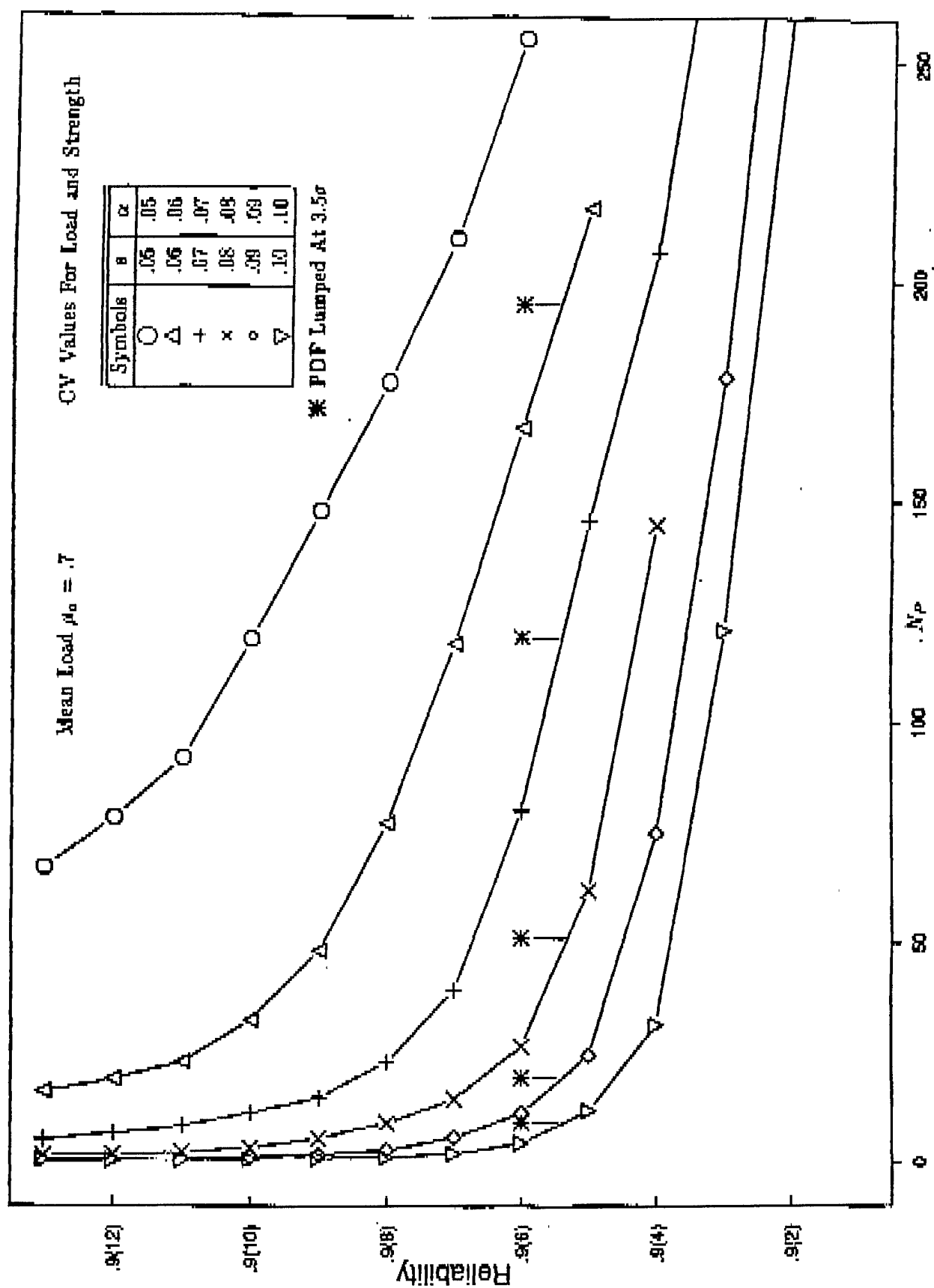
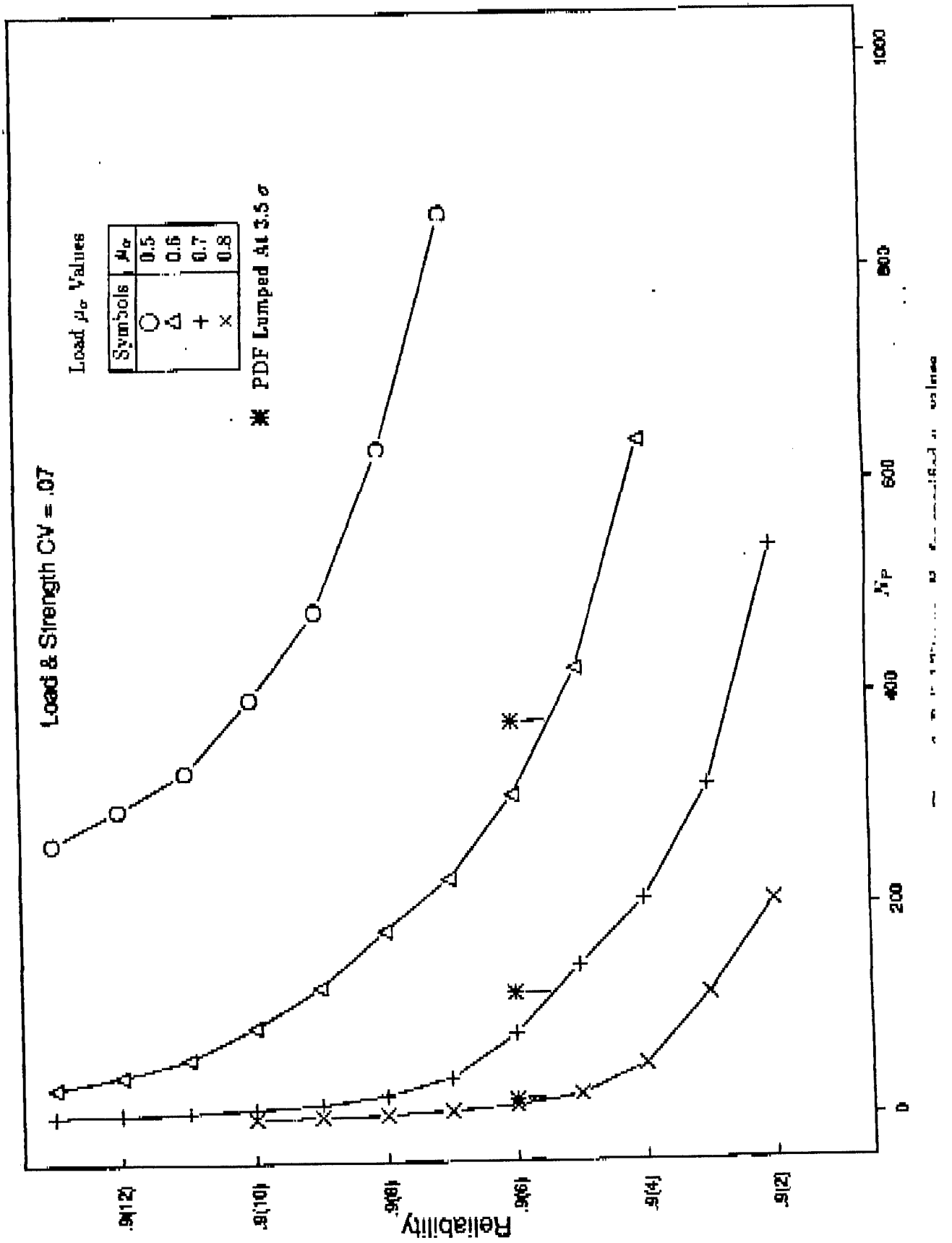


Figure 5. Reliability vs.  $N_p$  for specified CV values





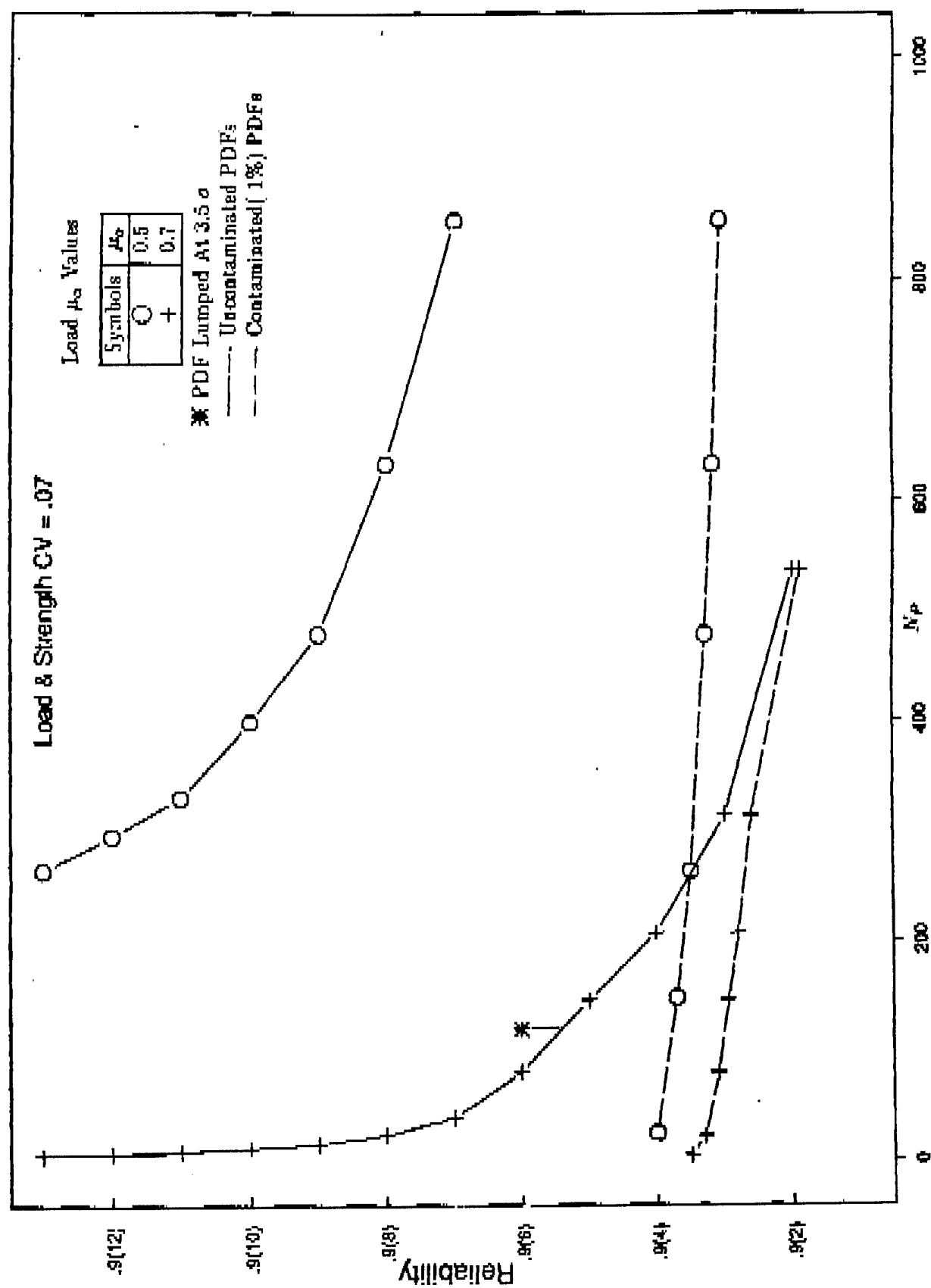


Figure 7. Reliability vs.  $N_p$ : Contaminated-Uncontaminated PDFs

# **Wavelets and Nonparametric Function Estimation: A Function Analytic Approach<sup>1</sup>**

Edward J. Wegman<sup>2</sup>  
Center for Computational Statistics  
George Mason University  
Fairfax, VA 22030

---

<sup>1</sup>This research was supported by the Army Research Office under Contract DAAL03-91-G-0039, the Office of Naval Research under Grant N00014-92-J-1303, and the National Science Foundation under Grant DMS9002237.

<sup>2</sup>Dr. Wegman is the Bernard J. Dunn Professor of Information Technology and Applied Statistics at George Mason University. This paper was presented as an invited paper at the 1991 Army Design of Experiments Conference held in Vicksburg, Mississippi.

This paper was presented at the Thirty-Seventh Conference in this series.

## Wavelets and Nonparametric Function Estimation: A Function Analytic Approach

**Abstract:** The problem of nonparametric function estimation has received a substantial amount of attention in the statistical literature over the last 15 years. To a very large extent, the literature has described kernel-based convolution smoothing solutions to the problems of probability density estimation and nonlinear regression. Among the sub-cultures within this literature has been a substantial effort at smoothing spline solutions. In the present paper, we discuss a general function analytic formulation of the problem. We show that a basis which spans  $L_2(\mathbb{R})$  can be used as a tool for constructing computational algorithms for optimal solutions to the generalized nonparametric function estimation problem. In particular wavelets form a doubly indexed set of basis functions for  $L_2(\mathbb{R})$  and may be used for computing optimal nonparametric function estimates. We discuss the basic theory of wavelets, and discuss connections of wavelets with multiresolution analysis, sub-band coding and conventional spectral analysis. We demonstrate the construction of compactly supported wavelets and illustrate the fractal character of a simple wavelet. We conclude our paper with a discussion of the relationship of wavelets to nonparametric function estimation, to time series analysis, to signal processing and to fractal geometry.

# Wavelets and Nonparametric Function Estimation: A Function Analytic Approach

## 1. Introduction.

Wavelets have captured the enthusiasm and imagination of many applied mathematicians and engineers both because of their important applications in signal and image processing and other engineering applications and also because of the inherent elegance of the techniques. Wavelets are described in detail in a number of locations. Much of the fundamental work was done by Daubechies and is reported in Daubechies, Grossman and Meyer (1986) and Daubechies (1988). Heil and Walnut (1989) provide a survey from a mathematical perspective while Rioul and Vetterli (1991) provide a survey from a more engineering perspective. The new book by Chui (1992) is an excellent integrated treatment which I believe is more mathematically sophisticated than the author supposes. In spite of its title as an introduction, it requires somewhat more mathematical depth and maturity and is best regarded as more of a monograph.

This present paper describes the basic wavelet theory in the context of the general statistical problem of nonparametric function estimation. Wegman (1984) describes a basic framework for optimal nonparametric function estimation. This framework captures the optimal estimation of a wide variety of practical function estimation problems in a common theoretical construct. Wegman (1984), however, only discusses the existence of such optimal estimators. In the present paper, we are interested in combining this optimality framework with more general wavelet and frame algorithms as computational devices for general optimal nonparametric function estimation. A new application of optimal nonparametric function estimation is found in Le and Wegman (1991).

In section 2, we discuss the optimal nonparametric function estimation framework. In section 3, we turn to a discussion of the general function analytic framework which leads to bases and frames. Section 4 introduces the notion of a wavelet basis and demonstrates the connection with Fourier series and Parseval's Theorem. In section 5 we turn to a spectral interpretation and show how the signal processing connection may be exploited to construct scaling functions and wavelets.

Finally, in section 6, we provide a synthesis of the connections among the varied elements of nonparametric function estimation, functional analysis, wavelets, time series, and signal processing.

## 2. Optimal Nonparametric Function Estimation.

Consider a general function,  $f(x)$ , to be estimated based on some sampled data, say  $x_1, x_2, \dots, x_n$ . This is, in fact, the most elementary estimation problem in statistical inference. Often the function,  $f$ , in question is the probability distribution function or the probability density function and most frequently the approach taken is to place the function within a parametric family indexed by some parameter, say  $\theta$ . Rather than estimate  $f$  directly, the parameter  $\theta$  is estimated with  $f_\theta$  then being estimated by  $\hat{f}_\theta = f_{\hat{\theta}}$ . Under a variety of circumstances, it is much more desirable to take a nonparametric approach so as to avoid problems associated with misspecification of parametric family. This is particularly the case when data is relatively plentiful and the information captured by the parametric model is not needed for statistical efficiency.

Probability density estimation and nonparametric, nonlinear regression are probably the two most widely studied nonparametric function estimation problems. However, other problems of interest which immediately come to mind are spectral density estimation, transfer function estimation, impulse response function estimation, all in the time series setting, and failure rate function estimation and survival function estimation in the reliability/biometry setting. While it may be the case that we simply may want an unconstrained estimate of the function, it is more often the case that we wish to impose one or more constraints, for example, positivity, smoothness, isotonicity, convexity, transience and fixed discontinuities to name a few appropriate constraints. By far, the most common assumption is smoothness and frequently the estimation is via a kernel or convolution smoother. We would like to formulate an optimal nonparametric framework.

We formulate the optimization problem as follows. Let  $\mathcal{H}$  be a Hilbert space of functions over  $\mathbb{R}$ , the real numbers (or  $\mathbb{C}$ , the complex numbers). For purposes of the present paper, we assume  $\mathbb{R}$  rather than  $\mathbb{C}$  unless otherwise specified. The techniques we outline here are not limited to a discussion of  $L_2(\mathbb{R})$  although quite often we do take  $\mathcal{H}$  to be  $L_2$ . In this case, we take

$$\langle f, g \rangle = \int f(x) g(x) d\mu(x),$$

is the familiar cubic spline.

The basic idea is to construct  $S \subseteq \mathcal{H}$  where  $S$  is the collection of functions,  $g$ , which satisfy our desired constraints such as smoothness or isotonicity. We wish to optimize  $L(g)$  over  $S$ . The optimized estimator will be an element of  $S$  and hence will inherit whatever properties we choose for  $S$ . The estimator will optimize  $L(g)$  and hence will be chosen according to whatever optimization criterion appeals to the investigator. In this sense we can construct designer estimators, i.e. estimators that are designed by the investigator to suit the specifics of the problem at hand.

Of course, in a wide variety of rather disparate contexts, many of these estimators are already known. However, they may be proven to exist in a general framework according to the following theorem.

**Theorem 2.1:**

Consider the following optimization problem:

Minimize (maximize)  $L(f)$  subject to  $f \in S \subseteq \mathcal{H}$ .

Then

- a) If  $\mathcal{H}$  is finite dimensional,  $L$  is continuous and convex (concave) and  $S$  is closed and bounded, then there exists at least one solution.
- b) If  $\mathcal{H}$  is infinite dimensional,  $L$  is continuous and convex (concave) and  $S$  is closed, bounded and convex, then there exists at least one solution.
- c) If  $L$  in a. or b. is strictly convex (concave), the solution is unique.
- d) If  $\mathcal{H}$  is infinite dimensional,  $L$  is continuous and uniformly convex (concave) and  $S$  is closed and convex, then there exists a unique solution.

**Proof:** A full proof is given in Wegman (1984). For completeness, we outline the basic elements here. a) For the finite dimensional case,  $S$  closed and bounded implies that  $S$  is compact. Choose  $f_n \in S$  such that  $L(f_n)$  converges to  $\inf\{L(f): f \in S\}$ . Because of compactness, there is a convergent subsequence  $f_{n_k}$  having a limit, say  $f_*$ . By continuity of  $L$

$$L(f_*) = \lim_{k \rightarrow \infty} L(f_{n_k}) = \inf\{L(f): f \in S\}.$$

$f_*$  is the required optimizer. For part b), we have the same basic idea except that  $S$  closed, bounded and convex implies that  $S$  is weakly compact. We use the weak continuity of  $L$ . Uniqueness follows by supposing both  $f_*$  and  $f_{**}$  are both minimizers.

where  $\mu$  is Lebesgue measure. We emphasize that this is not absolutely required. As usual  $\|f\| = \sqrt{\langle f, f \rangle}$ . A functional  $L: \mathcal{H} \rightarrow \mathbb{R}$  is *linear* if

$$L(\alpha f + \beta g) = \alpha L(f) + \beta L(g), \text{ for every } f, g \in \mathcal{H} \text{ and } \alpha, \beta \in \mathbb{R}.$$

$L$  is *convex* on  $S \subseteq \mathcal{H}$  if

$$L(tf + (1-t)g) \leq tL(f) + (1-t)L(g), \text{ for every } f, g \in S \text{ with } 0 \leq t \leq 1.$$

$L$  is *concave* if the inequality is reversed.  $L$  is *strictly convex (concave)* on  $S$  if the inequality is strict.  $L$  is *uniformly convex* on  $S$  if

$$tL(f) + (1-t)L(g) - L(tf + (1-t)g) \geq ct(1-t)\|f-g\|^2$$

for every  $f, g \in S$  and  $0 \leq t \leq 1$ .

We wish to use  $L$  as the general objective functional in our optimization framework. For example, if we are concerned with likelihood, we may consider the log likelihood,

$$L(f) = \sum_{i=1}^n \log f(x_i), \text{ } x_i \text{ are a random sample from } f.$$

If we have censored samples we may wish to consider

$$L(g) = \sum_{i=1}^n \delta_i \log g(x_i) + \sum_{i=1}^n (1 - \delta_i) \log \bar{G}(x_i),$$

$x_i$  again a random sample,  $\delta_i$  a censoring random variable,  $\bar{G} = 1 - G$ , and

$G(x) = \int_{-\infty}^x g(u) du$ . This is the censored log likelihood. Another example is the penalized least squares. In this case

$$L(g) = \sum_{i=1}^n (y_i - g(x_i))^2 + \lambda \int_a^b (Lg(u))^2 du.$$

Here  $L$  is a differential operator and the solution of this optimization problem over appropriate spaces is called a penalized smoothing  $L$ -spline. If  $L = D^2$ , then the solution

Then

$$L(tf_* + (1-t)f_{**}) < tL(f_*) + (1-t)L(f_{**}) = \inf\{L(f): f \in S\}.$$

This implies that neither  $f_*$  nor  $f_{**}$  is a minimizer which is a contradiction.  $\square$

This theorem gives us unified framework for the construction of optimal nonparametric function estimators. It does not, however, give us a definitive method for construction of nonparametric function estimators. We give a constructive framework in the next several sections. In closing this section we refer the reader to Wegman (1984) for the complete proof of Theorem 2.1 and many more examples of the use of this result.

### 3. Bases and Subspaces.

In this section, we discuss the basic theory of spanning bases and their application to function estimation. Consider  $f, g \in \mathcal{H}$ .  $f$  is said to be *orthogonal* to  $g$  written  $f \perp g$  if  $\langle f, g \rangle = 0$ . An element  $f$  is *normal* if  $\|f\| = 1$ . A family of elements, say  $\{e_\lambda: \lambda \in \Lambda\}$  is *orthonormal* if each element is normal and if for any pair  $e_1, e_2$  in the family,  $e_1 \perp e_2$ . A family  $\{e_\lambda: \lambda \in \Lambda\}$  is *complete* in  $S \subseteq \mathcal{H}$  if the only element in  $S$  which is orthogonal to every  $e_\lambda, \lambda \in \Lambda$  is 0. A *basis* or *base* of  $S$  is a complete orthonormal family in  $S$ . A Hilbert space has a countable basis if and only if it is separable, i.e. if and only if it has a countable dense subset. Ordinary  $L_p$  spaces are separable. We are now in a position to state the basic result characterizing bases of Hilbert spaces or subspaces. We write  $\text{span}(\{e_\lambda\})$  to be the minimal subspace containing  $\{e_\lambda\}$ . This is the space generated by the elements  $\{e_\lambda\}$ .

#### Theorem 3.1:

Let  $\mathcal{H}$  be a separable Hilbert space. If  $\{e_k\}_{k=1}^\infty$  is an orthonormal family in  $\mathcal{H}$ , then the following are equivalent.

- $\{e_k\}_{k=1}^\infty$  is a basis for  $\mathcal{H}$ .
- If  $f \in \mathcal{H}$  and  $f \perp e_k$  for every  $k$ , then  $f = 0$ .
- If  $f \in \mathcal{H}$ , then  $f = \sum_{k=1}^\infty \langle f, e_k \rangle e_k$ . (orthogonal series expansion)
- If  $f, g \in \mathcal{H}$ , then  $\langle f, g \rangle = \sum_{k=1}^\infty \langle f, e_k \rangle \langle g, e_k \rangle$ .
- If  $f \in \mathcal{H}$ ,  $\|f\|^2 = \sum_{k=1}^\infty |\langle f, e_k \rangle|^2$ . (Parseval's Theorem)

**Proof:**

a  $\Rightarrow$  b: Trivial by definition.



$b \Rightarrow c$ : We claim  $\mathcal{H} = \text{span}(\{e_k\})$ . If not there is  $f \neq 0$ ,  $f \in \mathcal{H}$  such that  $f \notin \text{span}(\{e_k\})$ . This implies that  $f \perp e_k$  for every  $k$ . But  $f \perp e_k$  for every  $k$  and  $f \neq 0$  is a contradiction to the  $\{e_k\}$  being a basis. Let  $\mathcal{H}_k = \text{span}(e_k)$ . Then  $\mathcal{H} = \text{span}(\bigcup_{k=1}^{\infty} \mathcal{H}_k) = \sum_k \mathcal{H}_k$ . This implies that for  $f \in \mathcal{H}$ ,

$$(3.1) \quad f = \sum_{k=1}^{\infty} c_k e_k.$$

Substituting (3.1) in the expression for the inner product yields

$$\langle f, e_j \rangle = \langle \sum_k c_k e_k, e_j \rangle = \sum_{k=1}^{\infty} c_k \langle e_k, e_j \rangle.$$

By the orthonormal property,  $\langle e_k, e_j \rangle = 1$ , if  $k=j$  and  $=0$ , otherwise. It follows that  $\langle f, e_j \rangle = c_j$ . Thus

$$(3.2) \quad f = \sum_{k=1}^{\infty} \langle f, e_k \rangle e_k.$$

$$c \Rightarrow d: \langle f, g \rangle = \langle f, \sum_{k=1}^{\infty} \langle g, e_k \rangle e_k \rangle = \sum_{k=1}^{\infty} \langle g, e_k \rangle \langle f, e_k \rangle.$$

$d \Rightarrow e$ : Let  $f = g$  in part d.

$e \Rightarrow a$ : If  $f \in \mathcal{H}$  and  $f \perp e_k$  for every  $k$  implies  $\langle f, e_k \rangle = 0$  for every  $k$ . This in turn implies that  $\|f\| = 0$ . Thus  $f = 0$ . This finally implies  $\{e_k\}_k$  is a basis.  $\square$

Thus given any basis  $\{e_k\}_k$ , we can exactly write  $f = \sum_{k=1}^{\infty} c_k e_k$  and we can estimate  $f$  by  $\sum_{k=1}^N \hat{c}_k e_k$ . Thus a computational algorithm for the optimal nonparametric function estimator can be based on this result from Theorem 3.1.c. However, this does not yet take into account the "design" set,  $S$ . In order to more carefully study the structure of  $S$  we consider the following result. In the following discussion let  $S \subseteq \mathcal{H}$ . Then define  $S^{\perp} = \{f \in \mathcal{H}: f \perp S\}$ .

**Theorem 3.2:**

If  $S \subseteq \mathcal{H}$  is a subset of  $\mathcal{H}$ , then

- $S^{\perp}$  is a subspace of  $\mathcal{H}$  and  $S \cap S^{\perp} \subseteq \{0\}$
- $S \subseteq S^{\perp\perp} = \text{span}(S)$
- $S$  is a subspace if and only if  $S = S^{\perp\perp}$ .

**Proof:**  $S^\perp$  is a linear manifold. To see this if  $f_1, f_2 \in S^\perp$ , then for every  $g \in S$ ,  $\langle a_1 f_1 + a_2 f_2, g \rangle = a_1 \langle f_1, g \rangle + a_2 \langle f_2, g \rangle = a_1 \cdot 0 + a_2 \cdot 0 = 0$ . Thus  $a_1 f_1 + a_2 f_2 \in S^\perp$ . This implies  $S^\perp$  is a linear manifold which is sufficient to show that  $S^\perp$  is a subspace provided we can show  $S^\perp$  is closed. To see this if  $f \in \text{closure}(S^\perp)$ , then there exists  $\{f_n\} \subseteq S^\perp$  such that  $f = \lim_{n \rightarrow \infty} f_n$  and for every  $g \in S$ ,  $\langle f_n, g \rangle = 0$ . But  $\langle f, g \rangle = \lim_{n \rightarrow \infty} \langle f_n, g \rangle = \lim_{n \rightarrow \infty} 0 = 0$ . This implies  $f \perp S$  which in turn implies  $f \in S^\perp$ . Part b follows from part a by replacing  $S$  by  $S^\perp$ . Part c is straightforward application of the two previous parts.  $\square$

Suppose now that we have a basis for  $\mathcal{H}$ , call it  $\{e_k\}_{k=1}^\infty$ . This basis obviously also spans subset  $S$  of  $\mathcal{H}$  and hence any of our "designer" functions in  $S$  can be written in terms of the basis,  $\{e_k\}_{k=1}^\infty$ . The unnecessary basis elements will simply have coefficients of 0. In a sense, however, this basis is too rich and in a noisy estimation setting superfluous basis elements will only contribute to estimating noise. As part of our "designer" set,  $S$ , philosophy, we would like to have a minimal basis set for  $S$ . Theorem 3.2 gives us a test for this condition. Consider a basis  $\{e_k\}_{k=1}^\infty$  for  $\mathcal{H}$ . Form  $B_S$  which is to be a basis for  $S$ . We define  $B_S$  by the following routine. If there is a  $g \in S$  such that  $\langle g, e_k \rangle \neq 0$ , then let  $e_k \in B_S$ . If on the other hand there is a  $g \in S^\perp$  such that  $\langle g, e_k \rangle \neq 0$ , then let  $e_k \in B_{S^\perp}$ . Unfortunately, it may not be that  $B_S \cap B_{S^\perp} = \emptyset$ . But this algorithm yields  $\{e_k\} = B_S \cup B_{S^\perp}$ . Moreover  $S \subseteq \text{span}(B_S)$ . Thus we may be able to eliminate unnecessary basis elements. We may also be able to re-normalize the basis elements using a Gram-Schmidt orthogonalization procedure to make  $B_S \perp B_{S^\perp}$ . Usually if we know the properties of the set,  $S$ , we desire and the nature of the basis set  $\{e_k\}$ , it will be straightforward to construct a test function,  $g$ , with which to construct the basis set,  $B_S$ . If  $S$  is a subspace, then  $S = \text{span}(B_S)$ . In any case we can carry out our estimation by

$$(3.3) \quad \hat{f} = \sum_{e_k \in B_S} \hat{c}_k e_k.$$

In a completely noiseless setting (3.1) is really an equality in norm, i.e.  $\|f - \sum_k c_k e_k\| = 0$ . If  $\mathcal{H}$  is  $L_2(\mu)$ , with  $\mu$  Lebesgue measure, then (3.1) is really

$$(3.4) \quad f = \sum_k c_k e_k, \text{ almost everywhere } \mu \text{ with } c_k = \langle f, e_k \rangle.$$

This choice of  $c_k$  is a minimum norm choice. However, in a noisy setting, i.e. where we

do not know  $f$  exactly, we cannot compute  $c_k$  directly. However, we may be able to estimate  $c_k$  by standard inference techniques.

**Example 3.1. Norm Estimate.** The minimum norm estimate of  $c_k$  is the choice which minimizes  $\|f - \sum_k c_k e_k\|$ , i.e.  $c_k = \langle f, e_k \rangle$ . In the  $L_2$  context,

$$\langle f, e_k \rangle = \int_{\mathbb{R}} f(x) e_k(x) d\mu(x).$$

If  $f$  is a probability density function, then  $\langle f, e_k \rangle = E[e_k]$  which can simply be estimated by  $n^{-1} \sum_{j=1}^n e_k(x_j)$ , where  $x_j, j = 1, \dots, n$  is the sample of observations.

**Example 3.2. General Form of Estimate.** In the general context with optimization functional  $\mathcal{L}$  we have

$$(3.5) \quad \mathcal{L}(f) = \mathcal{L}\left(\sum_{c_k \in B_s} c_k e_k\right) \triangleq \mathcal{L}(\{c_k\}).$$

Since (3.5) is a function of a countable number of variables,  $\{c_k\}$ , we can find the normal equations and with the appropriate choice of basis, find a solution. For this we will typically assume  $\mathcal{L}$  is twice differentiable with respect to all  $c_k$ . A wide variety of bases have been studied. These include Laguerre polynomials, Hermite polynomials and other orthonormal systems. Perhaps the most well-known orthonormal system is the fundamental sinusoids which span  $L_2(0, 2\pi)$ . From the title and theme of this paper, one might reasonable guess that wavelets form another orthogonal system. We discuss the connection in the next section.

## 4. Fourier Analysis and Wavelets.

### 4.1 Bases for $L_2(0, 2\pi)$ .

Let us consider the set of square-integrable functions on  $(0, 2\pi)$  which we denote by  $L_2(0, 2\pi)$ .  $L_2(0, 2\pi)$  is a Hilbert space and a traditional choice of an orthonormal basis for this space has been  $e_k(x) = e^{ikx}$ , the complex sinusoids. Thus any  $f$  in  $L_2(0, 2\pi)$  has the Fourier representation by Theorem 3.1.c

$$f(x) = \sum_{k=-\infty}^{\infty} c_k e^{ikx}$$

where the constants  $c_k$  are the Fourier coefficients defined by

$$c_k = \frac{1}{2\pi} \int_0^{2\pi} f(x) e^{-ikx} dx.$$

This pair of equations represent the discrete Fourier transform and the inverse Fourier transform and is the foundation of harmonic analysis. An interesting feature of this complex sinusoids as a base for  $L_2(0, 2\pi)$  is that  $e_k(x) = e^{ikx}$  can be generated from the superpositions of dilations of a single function,  $e(x) = e^{ix}$ . By this we mean that

$$e_k(x) = e(kx), \quad k = \dots, -1, 0, 1, \dots$$

These are *integral dilations* in the sense that  $k \in J$ , the integers. The concept of dilations of a fixed generating function is central to the formation of wavelet bases as we shall see shortly.

A well known consequence of Theorem 3.1.e for the complex sinusoid basis is the Parseval Theorem. For this base, we have

**Theorem 4.1:** (Parseval's Theorem):

$$(4.1) \quad \|f\|^2 = \int_0^{2\pi} |f(x)|^2 dx = \sum_{k=-\infty}^{\infty} |c_k|^2.$$

Equation (4.1) is known as Parseval's Theorem in harmonic analysis and states that the square norm in the frequency domain is equal to the square norm in the time domain.

While the space  $L_2(0, 2\pi)$  is an extremely useful one, for general problems in nonparametric function estimation we are much more interested in  $L_2(\mathbb{R})$ . We can think of  $L_2(0, 2\pi)$  as with functions on the finite support  $(0, 2\pi)$  or as periodic functions on  $\mathbb{R}$ . In the latter case it is clear that the infinitely periodic functions of  $L_2(0, 2\pi)$  and the square integrable functions of  $L_2(\mathbb{R})$  are very different. In the latter case the function,  $f(x) \in L_2(\mathbb{R})$ , must converge to 0 as  $x \rightarrow \pm\infty$ . The generating function  $e(x) = e^{ix}$  clearly does not have that behavior and is inappropriate as a basis generating function for  $L_2(\mathbb{R})$ . What is needed is a generating function,  $e(x)$ , which also has the property that  $e(x) \rightarrow 0$  as  $x \rightarrow \pm\infty$ . Thus we want to generate a basis from a function which will decay to 0 relatively rapidly, i.e. we want little waves or *wavelets*.

## 4.2 Wavelet Bases.

Let us begin by considering a generating function  $\psi$  which we will think of as our *mother wavelet* or basic wavelet. The idea is that, just as with the sinusoids, we wish to consider a superposition of dilations of the basic waveform  $\psi$ . For technical convergence reasons which we shall explain later we wish to consider dyadic dilations rather than simply integral translations. Thus for the first pass, we are inclined to consider  $e_j(x) = 2^{j/2}\psi(2^j x)$ . Unfortunately, because of the decay of  $\psi$  to 0 as  $x \rightarrow \pm\infty$ , the elements  $\{e_j\}$  are not sufficient to be a basis for  $L_2(\mathbb{R})$ . We accommodate this by adding translates to get the doubly indexed functions  $e_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$ . We choose  $\psi$  such that

$$\int_{\mathbb{R}} \frac{|\hat{\psi}(\omega)|^2}{\omega} d\omega \text{ exists.}$$

Here  $\hat{\psi}$  is the Fourier transform of  $\psi$ . Under certain choices of  $\psi$ ,  $e_{j,k}$  forms a doubly indexed orthonormal basis for  $L_2$  (actually also for Sobolev spaces of higher order as well). As we shall see in the next section, a wavelet basis due to the dilation-translation nature of its basis elements admits an interpretation of a simultaneous time-frequency decomposition of  $f$ . Moreover using wavelets, fewer basis elements are required for fitting sharp changes or discontinuities. This implies faster convergence in "non-smooth" situations by the introduction of "localized" basis elements.

**Example 3.1 Continued:** Notice that

$$c_{j,k} = \langle f, e_{j,k} \rangle = \int_{-\infty}^{\infty} 2^{j/2} \psi(2^j x - k) f(x) dx.$$

In the density estimation case

$$c_{j,k} = E \left( 2^{j/2} \psi(2^j x - k) \right).$$

Thus a natural estimator is

$$\hat{c}_{j,k} = \frac{2^{j/2}}{n} \sum_{i=1}^n \psi(2^j x_i - k),$$

where  $x_i$ ,  $i = 1, \dots, n$  is the set of observations.

Notice that we can construct a Parseval's Theorem for Wavelets.

**Theorem 4.2: (Parseval's Theorem for Wavelets)**

$$(4.2) \quad \|f\|^2 = \int_{-\infty}^{\infty} |f(x)|^2 dx = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} |c_{j,k}|^2 = \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} |c_{j,k}|^2.$$

We shall give additional interpretation to this equation in the next section.

### 4.3 Frames.

Frames were originally introduced by Duffin and Schaeffer (1952) and have become the subject of increased interest with the emergence of the interest in wavelets. Frames are a generalization of bases, but share many of the same series representation properties of bases. Let  $\{e_k\}_{k=1}^{\infty}$  be a collection of elements in  $\mathcal{H}$ . The collection  $\{e_k\}_{k=1}^{\infty}$  is a *frame* if there are positive numbers  $A, B$  such that

$$A \|f\|^2 \leq \sum_{k=1}^{\infty} |\langle f, e_k \rangle|^2 \leq B \|f\|^2.$$

$A$  and  $B$  are called the *frame bounds*. If  $A = B$ , the frame is said to be *tight*. If no  $e_k$  can be dropped from the frame, then the frame is said to be *exact*. Notice that if  $A = B = 1$  and the frame is exact, then we have

$$\|f\|^2 = \sum_{k=1}^{\infty} |\langle f, e_k \rangle|^2.$$

By Theorem 3.2, the frame is then a basis. Frames are, in general, not bases, but as indicated earlier they share some of the same properties. In particular, if we define an operator  $T$  by  $Tf = \sum_{k=1}^{\infty} \langle f, e_k \rangle e_k$ , then  $T$  is called the *frame operator*. It is easy to verify that  $A \langle f, f \rangle \leq \langle Tf, f \rangle \leq B \langle f, f \rangle$ .  $T$  is invertible and if  $T^{-1}$  is the inverse operator, then  $\{T^{-1}(e_k)\}_{k=1}^{\infty}$  is also a frame, called the *dual frame* with frame bounds  $1/A$  and  $1/B$ . In particular,

$$\sum_{k=1}^{\infty} |\langle f, T^{-1}(e_k) \rangle|^2 \leq \frac{1}{A} \|f\|^2.$$

If the frame is exact, the sequences  $\{e_k\}_{k=1}^{\infty}$  and  $\{T^{-1}(e_k)\}_{k=1}^{\infty}$  are *biorthonormal*, that is  $\langle e_j, T^{-1}(e_k) \rangle = \delta_{jk}$  where  $\delta_{jk} = 1$  if  $j = k$  and  $\delta_{jk} = 0$  otherwise. Most

importantly,

$$(4.3) \quad f = \sum_{k=1}^{\infty} \langle f, e_k \rangle T^{-1}(e_k)$$

or

$$(4.4) \quad f = \sum_{k=1}^{\infty} \langle f, T^{-1}(e_k) \rangle e_k.$$

The equality (4.4) is a particularly a useful form since it so closely parallels (3.2) which is the fundamental method for constructing function estimators with basis functions.

It is perhaps useful to point out the utility of frames. We see two particularly useful settings. First, if we wish to take the union of a finite number of admissible spaces, say  $S_i$ . Then the union of the bases,  $\cup_i B_{S_i}$ , is a frame for  $\cup_i S_i$ . Secondly, if we have an admissible space  $S$  with basis  $B_S$ , but we wish to add a few additional elements, say  $\{g_j, j \in \mathbb{Z}\}$ . Then  $B \cup \{g_j, j \in \mathbb{Z}\}$  is a frame for the enlarged space. This is useful, for example, if we know there are some discontinuities in an otherwise smooth (e.g. Sobolev space) space.

We conclude this section by noting that it is straightforward to show that for  $f, g \in \mathcal{H}$ ,

$$\langle f, g \rangle = \sum_{k=1}^{\infty} \langle f, e_k \rangle \langle g, T^{-1}(e_k) \rangle = \sum_{k=1}^{\infty} \langle f, T^{-1}(e_k) \rangle \langle g, e_k \rangle$$

in analogy to Theorem 3.1.d. Letting  $f = g$  in the above, it follows immediately that

$$\|f\|^2 = \sum_{k=1}^{\infty} \langle f, e_k \rangle \langle f, T^{-1}(e_k) \rangle$$

which are frame analogs to the results of Theorem 3.1.

## 5. Spectral Interpretation of Wavelets.

### 5.1 Continuous Wavelet Transforms.

We have at this stage alluded to wavelets and frames, but have not really explained why a wavelet decomposition is of any particular interest in nonparametric function estimation. To appreciate this let us look again at the traditional methods of Fourier or harmonic analysis in statistics. Corresponding to a function  $f(x)$ , there is a

function

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} f(x) e^{-i\omega x} dx$$

which is the *Fourier transform* of  $f$ . The *inverse Fourier transform* can be computed by

$$(5.1) \quad f(x) = 2\pi \int_{-\infty}^{\infty} \hat{f}(\omega) e^{i\omega x} d\omega.$$

We may not always have a complete version of  $\hat{f}$  available so it is useful to have a sampled version. In this case,

$$(5.2) \quad f(x) = \sum_{k=-\infty}^{\infty} c_k e^{ikx}$$

which is the so-called discrete Fourier transform alluded to in section 3. Here

$$c_k = \frac{1}{2\pi} \int_0^{2\pi} f(x) e^{-ikx} dx$$

are the *Fourier coefficients*. The fast Fourier transform (FFT) is a fast computational algorithm for computing the discrete Fourier transform. Fourier methods are appropriate for analysis of stationary stochastic processes or time series since stationarity implies that covariance structure is invariant with time. Because the Fourier transform is usually applied to the covariance function to obtain the spectral density, the frequency structure of a stationary process is invariant with time. It is clear that traditional Fourier methods are not suitable for non-stationary or transient stochastic processes. Thus even though the fundamental sinusoids span  $L_2(0, 2\pi)$ , the series in (5.2) is not a parsimonious representation of  $f(x)$  and hence will be slow to converge in nonstationary settings.

It is desirable to localize in both time and frequency. One approach to localization in both time and frequency has been the *short term Fourier transform (STFT)* or as it is also known, the *Gabor transform* given by the expression below.

$$\hat{\hat{f}}(\omega, \tau) = \int_{-\infty}^{\infty} f(x) w^*(t - \tau) e^{-i\omega x} dx$$

where  $w^*$  is the complex conjugate of  $w$ . In other words the STFT is a windowed



Fourier transform. There are unfortunately faults with this idea. The STFT is poor at resolving wavelengths longer than the window width, that is, it is poor at resolving low frequencies. Conversely, the STFT is poor at localizing high frequencies because the window average energy over the window width. That is for fixed window width, the STFT time and frequency resolutions are limited by the Heisenberg inequality (time-bandwidth product bounded below by  $(4\pi)^{-1}$ ). What is needed is a scheme which allows for large window widths at low frequencies and very small window widths at high frequencies.

The basic wavelet idea is to use a transient waveform as in the STFT, but to increase time resolution by keeping a constant relative bandwidth as frequency increases. As we have seen, we choose a prototype wavelet,  $\psi(t)$ , and consider dilations and translations of this mother wavelet  $\psi$  which are the *affine wavelets*

$$\psi_{a\tau}(t) = \sqrt{a} \psi(a(t - \tau)).$$

The *continuous wavelet transform (CWT)* is defined by

$$\hat{\hat{f}}(\tau, a) = \int_{-\infty}^{\infty} f(x) \psi_{a,\tau}^*(x) dx$$

and the *inverse wavelet transform* is given by

$$(5.3) \quad f(x) = c \int \int_{a>0} \hat{\hat{f}}(\tau, a) \psi_{a,\tau}(x) \frac{d\tau da}{a^2}.$$

This latter equation is sometimes reparametrized with  $a = e^\nu$ . Just as we deal with a Fourier series representation of  $f$ , we would like to deal with a wavelet series representation. The parameter  $a$  (or its surrogate  $\nu$ ) is the *dilation parameter* and is the analog of the frequency,  $\omega$ , in the ordinary harmonic case. It is more properly thought of as a scale parameter. The parameter,  $\tau$ , is a time-location parameter.

As previously indicated, a decrease in scale corresponds to an increase in frequency (large scale for low frequencies, small scale for high frequencies). This basic concept is illustrated in Figure 5.1. Notice for a STFT, as the frequency increases the resolution stays fixed as illustrated by the grid in Figure 5.1.a., i.e. the scale stays the same and more cycles are included in the window for higher frequencies as indicated in Figure 5.1.c. This implies an inability to localize at high frequencies for the STFT.

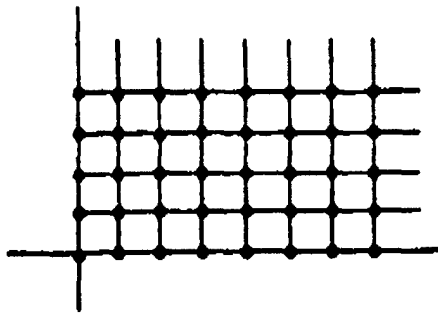


Figure 5.1.a Fixed-width grid for STFT.

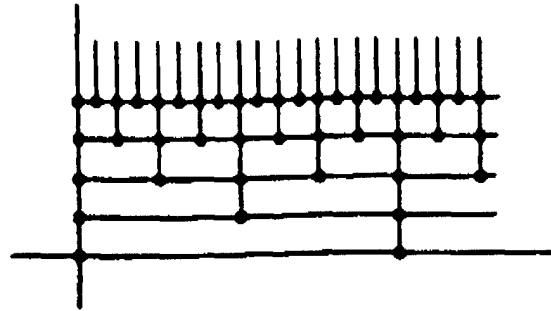


Figure 5.1.c Dyadic grid for wavelet transform.

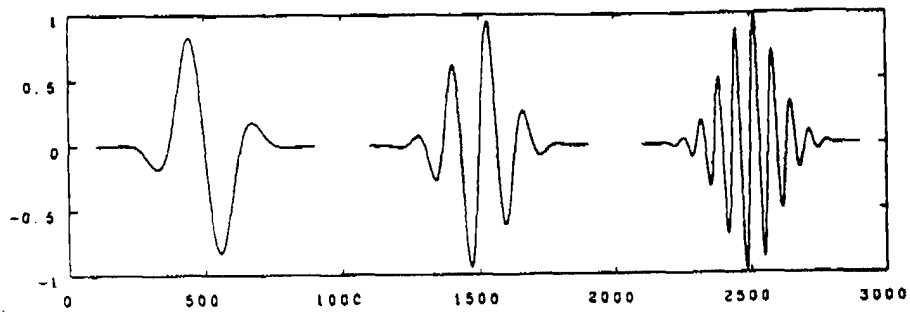


Figure 5.1.b Windowed waveforms for STFT. Window width is constant so more cycles appear at higher frequencies.

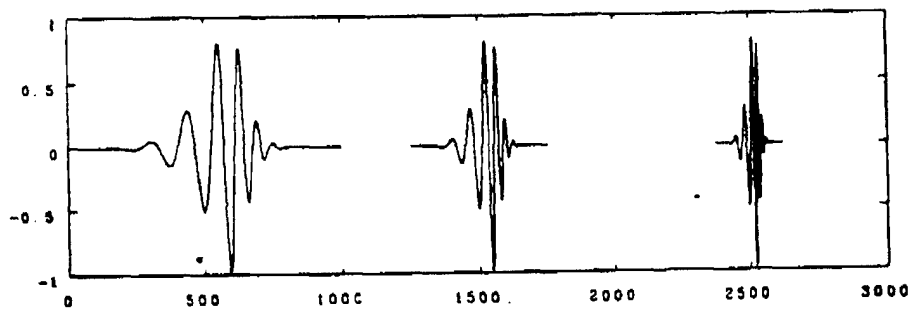


Figure 5.1.d Daubechies' 20-term FIR-based wavelet with several dilations. Dilation compresses or expands waveform, but does not change the number of oscillations.

Similarly at low frequencies, the window width stays constant so that only a fraction of a cycle may appear within the smoothing window. This implies an inability to resolve low frequencies. For the wavelet transform, however, has a waveform with a fixed basic structure,  $\psi$  which is dilated and translated, but otherwise unchanged as illustrated in Figure 5.1.d. This implies an increased time resolution at high frequencies as illustrated in Figure 5.1.b. Conversely since the scale increases at low frequencies, (i.e. the same number of oscillations are included as the waveform is expanded), low frequency resolution is also improved. We may discretize the time and scale parameters by  $a = a_0^j$  and  $\tau = k\tau_0/a_0^j$  where  $j$  and  $k$  are integers and the wavelets are

$$\psi_{jk}(t) = a_0^{j/2} \psi(a_0^j t - k\tau_0)$$

with wavelet coefficients given by

$$c_{jk} = \int_{-\infty}^{\infty} f(x) \psi_{jk}^*(x) dx.$$

The illustrations in Figure 5.1.b. are for dyadic discretizations, i.e.  $a_0 = 2$ . If  $a_0$ ,  $\tau_0$  and  $\psi(t)$  have the appropriate properties, we might expect as in the case of Fourier series that

$$(5.4) \quad f(x) = \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} c_{jk} \psi_{jk}(x).$$

This is precisely the form we had in section 4 associated with either an orthonormal basis or a frame. Thus if we can show that the wavelets either form a basis or a frame then the representation (5.4) will obtain. Notice that if  $a_0$  is arbitrarily close to one, then the double sum in equation (5.4) is a Riemann sum approximation to the double integral in (5.3) and hence it is reasonable to believe that such conditions are possible. To make wavelets computationally feasible, however, we would like to have  $a_0 = 2$ .

Thus we may reasonable ask under what conditions do wavelets form an orthonormal basis for  $L_2$  or indeed for other spaces. In addition we would like to have wavelets with compact support. If the support for the mother wavelet,  $\psi$ , is not bounded, then every term in the doubly infinite series will contribute to the value of  $f(x)$  and we will gain little computational advantage to using wavelets. In addition, we would like the wavelets to be dyadic, that is,  $a_0 = 2$  with  $\tau_0 = 1$ . These choices result in

no oversampling so that orthonormal bases are possible. Construction of wavelets under such conditions was done by Daubechies (1988) in a computation closely related to the idea of multi-resolution analysis.

## 5.2 Multi-Resolution Analysis.

To understand multi-resolution analysis let us first consider the construction of space  $W_j = \text{span}\{\psi_{j,k}: k \in J\}$ . That is we fix the dilation and consider the space generated by all possible translates. For purposes of discussion in this section we take  $a_0 = 2$  and  $r_0 = 1$  so that  $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$ . We may write  $L_2(\mathbb{R})$  as a direct sum of the  $W_j$ ,  $L_2(\mathbb{R}) = \sum_{j \in J} W_j$  so that any function  $f \in L_2(\mathbb{R})$  may be written as

$$f(x) = \dots + d_{-1}(x) + d_0(x) + d_1(x) + \dots$$

where  $d_j \in W_j$ . If  $\psi$  is an orthogonal wavelet, then  $W_j \perp W_k$ ,  $k \neq j$ . We shall assume  $\psi$  to be an orthogonal wavelet in what follows. Notice that as  $j$  increases, the basic wavelet form  $\psi(2^j x - k)$  contracts representing higher "frequencies." For each  $j$  we may consider the direct sum  $V_j$  given by:

$$V_j = \dots + W_{j-2} + W_{j-1} = \sum_{m=-\infty}^{j-1} W_m.$$

The  $V_j$  are closed subspaces and represent spaces of functions with all "frequencies" at or below a given level of resolution. The set of spaces  $\{V_j\}$  has the following properties:

- 1) They are nested in the sense that  $V_j \subseteq V_{j+1}$ ,  $j \in J$ .
- 2) Closure  $(\cup_{j \in J} V_j) = L_2(\mathbb{R})$ .
- 3)  $\cap_{j \in J} V_j = \{0\}$ .
- 4)  $V_{j+1} = V_j + W_j$ .
- 5)  $f(x) \in V_j$  if and only if  $f(2x) \in V_{j+1}$ ,  $j \in J$ .

1), 4) and 5) follow directly from the definition of  $V_j$ . 2) is a straightforward consequence of the fact that  $\cup_{j \in J} W_j = L_2(\mathbb{R})$ . 3) follows because of the orthogonality property.

Any  $f \in L_2(\mathbb{R})$  can be projected into  $V_j$ . As we have seen with  $j$  increasing the the "frequency" of the wavelet increases which can be interpreted as higher resolution.

Thus the projection,  $P_j f$ , of  $f$  into  $V_j$  is an increasingly higher resolution approximation to  $f$  as  $j \rightarrow \infty$ . Conversely, as  $j \rightarrow -\infty$ ,  $P_j f$  is an increasingly blurred (smoothed) approximation to  $f$ . We shall take  $V_0$  as the *reference subspace*. Suppose now that we can find a function  $\phi$  and that we can define  $\phi_{j,k}(x) = 2^{j/2} \phi(2^j x - k)$  such that

$$V_0 = \text{span}\{\phi_{0,k} : k \in J\}.$$

Then by property 5),  $V_j = \text{span}\{\phi_{j,k} : k \in J\}$ . While we began our discussion with the notion of wavelets and have seen some of the consequences, we could have actually begun a discussion with the function  $\phi$ .

**Definition.** A function  $\phi$  generates a *multiresolution analysis* if it generates a nested sequence of spaces having properties 1), 2), 3) and 5) such that  $\{\phi_{0,k}, k \in J\}$  forms a basis for  $V_0$ . If so, then  $\phi$  is called the *scaling function*.

For the final discussion of this section, let us consider a multiresolution analysis in which  $\{V_j\}$  are generated by a scaling function  $\phi \in L_2(\mathbb{R})$  and  $\{W_j\}$  are generated by a mother wavelet function  $\psi \in L_2(\mathbb{R})$ . Any function  $f \in L_2(\mathbb{R})$  can be approximated as closely as desired by  $f_m$  for some sufficiently large  $m \in J$ . Notice  $f_m = f_{m-1} + d_{m-1}$  where  $f_{m-1} \in V_{m-1}$  and  $d_{m-1} \in W_{m-1}$ . This process can be recursively applied say  $l$  times until we have  $f \cong f_m = d_{m-1} + d_{m-2} + \dots + d_{m-l} + f_{m-l}$ . Notice that  $f_{m-l}$  is a highly smoothed version of the function. Indeed, this suggests that a statistical procedure might be to form a highly smoothed (even overly smoothed) approximation to a function to be estimated. The sequence  $d_{m-l}$  through  $d_{m-1}$  form the higher resolution wavelet approximations. Many of the wavelet coefficients  $c_{m-i,k}$  used for constructing  $d_{m-i}$ ,  $i = 1, \dots, l$  are likely to be 0 and hence can contribute to a very parsimonious representation of the function  $f$ . Indeed, a wavelet decomposition is a natural suggestion for a technology for high definition television (HDTV). If  $f_{m-l}$  represents the lower resolution conventional NTSC TV signal, then to reconstruct a high resolution image all that is needed is the difference signal which could be parsimoniously represented by the wavelet coefficients  $c_{m-i,k}$ ,  $i = 1, \dots, l$  and  $k \in J$  most of which would be 0.

Most importantly, however, is the observation that the scaling function  $\phi \in V_0$  and the mother wavelet  $\psi \in W_0$  implies that both are in  $V_1$ . Since  $V_1$  is generated by

$\phi_{1,k}(x) = 2^{1/2} \phi(2x - k)$ , there are sequences  $\{g(k)\}$  and  $\{h(k)\}$  such that

$$(5.5) \quad \phi(x) = \sum_{k \in J} g(k) \phi(2x - k) \text{ and } \psi(x) = \sum_{k \in J} h(k) \phi(2x - k).$$

This remarkable result gives us a construction for the mother wavelet in terms of the scaling function. These equations are called the two-scale difference equations. We can give a time series interpretation to these equations. Lets consider an original discrete time function,  $f(n)$ , to which we apply the filter

$$y(n) = \sum_{k \in J} g(k) f(2n - k).$$

First of all we note that there is a scale change due to subsampling by two, i.e. a shift by two in  $f(n)$  results in a shift of one in  $y(n)$ . The scale of  $y$  is only half that of  $f$ . Otherwise this is a low pass filter with impulse response function  $g$ . Let us consider iterating this equation so that

$$(5.6) \quad y^{(j)}(n) = \sum_{k \in J} g(k) y^{(j-1)}(2n - k).$$

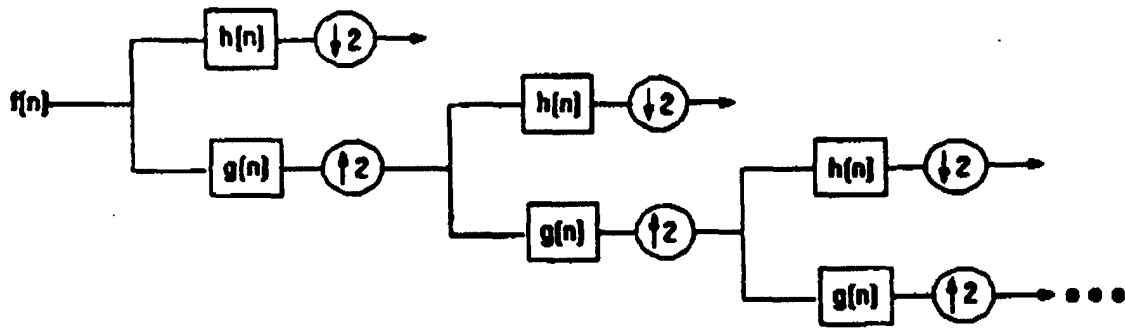


Figure 5.2 Decomposition scheme for multiresolution analysis.

Notice that if this procedure converges, it converges to a fixed point which will be  $\phi$ . This iterative procedure with repeated down sampling by two is illustrated in Figure 5.2 and is suggestive of a method for constructing wavelets. If  $g$  is a finite impulse response filter, the construction of a complementary high-pass filter is accomplished with a filter,  $h$ , whose impulse response is given by  $h(l-1-n) = (-1)^n g(n)$ . This scheme is called

*sub-band coding* in the electrical engineering literature. The low-pass band is given by

$$(5.7) \quad y_0(n) = \sum_{k \in J} g(k)f(2n-k)$$

while the high-pass band is given by

$$(5.8) \quad y_1(n) = \sum_{k \in J} h(k)f(2n-k).$$

The filter impulses as defined form an orthonormal set so that the  $f$  may be reconstructed by

$$(5.9) \quad f(n) = \sum_{k \in J} [y_0(k)g(2k-n) + y_1(k)h(2k-n)].$$

The sub-band coding scheme may be repeatedly applied to form the nested sequence as illustrated in Figure 5.2. The nested sequence of  $\{V_j\}$  is then essentially obtained by recursively downsampling and filtering a function with a low-pass filter whose impulse response function is  $g(\cdot)$ .

### 5.3 Construction of Scaling Functions and Mother Wavelets.

We have already hinted that the scaling function may be constructed as the fixed point of the down-sampled, low-passed filter equation (5.6). This can be formalized by considering what statisticians would call the generating function of  $g(n)$  and what electrical engineers call the  $z$ -transform of  $g(\cdot)$ .

$$(5.10) \quad G(z) = \frac{1}{2} \sum_{j \in J} g(j) z^j.$$

Notice if  $z = e^{-i\omega/2}$ , then (5.10) is essentially the Fourier transform of the impulse response function  $g(\cdot)$ . In this case, the first equation in (5.5) may be written as

$$(5.11) \quad \hat{\phi}(\omega) = G(z)\hat{\phi}\left(\frac{\omega}{2}\right), \text{ with } z = e^{-i\omega/2}.$$

This, of course, follows because the Fourier transform of a convolution is a product. This recursive equation may be iterated to obtain

$$(5.12) \quad \hat{\phi}(\omega) = \prod_{k=1}^{\infty} G(e^{-i\omega/2^k}) \hat{\phi}(0).$$

We may take  $\hat{\phi}$  to be continuous and  $\hat{\phi}(0) = 1$ . Based on (5.12) we may recover  $\phi(\cdot)$  and based on this result, the equation  $h(l-1-n) = (-1)^n g(n)$  and the second equation of (5.5) we may recover the mother wavelet,  $\psi(\cdot)$ . Thus Daubechies' original construction shows that wavelets with compact support can be based on finite impulse response filters which was originally motivated by multiresolution analysis. Theorem 5.1 below summarizes the Daubechies' result.

**Theorem 5.1: (Daubechies' Wavelet Construction):**

Let  $g(n)$  be a sequence such that

- a)  $\sum_{n \in J} |g(n)| |n|^\epsilon < \infty$  for some  $\epsilon > 0$ ,
- b)  $\sum_{n \in J} g(n-2j) g(n-2k) = \delta_{jk}$ ,
- c)  $\sum_{n \in J} g(n) = 1$ .

Suppose that  $\hat{g}(\omega) = G(e^{-i\omega/2}) = 2^{-1/2} \sum_{n \in J} g(n) e^{-in\omega/2}$  can be written as

$$\hat{g}(\omega) = \left[ \frac{1}{2} (1 + \epsilon^{-i\omega/2})^N \right] \cdot \left[ \sum_{n \in J} f(n) e^{in\omega} \right]$$

where

- d)  $\sum_{n \in J} |f(n)| |n|^\epsilon < \infty$  for some  $\epsilon > 0$
- e)  $\sup_{\omega \in \mathbb{R}} \left| \sum_{n \in J} f(n) e^{in\omega} \right| < 2^{N-1}$ .

Define

$$h(n) = (-1)^n g(-n+1),$$

$$\hat{\phi}(\omega) = \prod_{k=1}^{\infty} G(e^{-i\omega/2^k}),$$

$$\psi(x) = \sum_{k \in J} h(k) \phi(2x - k).$$

Then the orthonormal wavelet basis is  $\psi_{jk}$  determined by the mother wavelet  $\psi$ . Moreover, if  $g(n) = 0$  for  $n > n_0$ , then the wavelets so determined have compact support.



We state this result without proof. We note that Daubechies also shows that the mother wavelet,  $\psi$ , cannot be an even function and also have a compact support. The exception to this is the trivial constant function which gives rise to the so-called Haar basis. Daubechies illustrates this computation with the example of  $g$  given by  $g(0) = (1 + \sqrt{3})/8$ ,  $g(1) = (3 + \sqrt{3})/8$ ,  $g(2) = (3 - \sqrt{3})/8$  and, finally,  $g(3) = (1 - \sqrt{3})/8$ . This wavelet is illustrated in Figure 5.3.

## 6. Conclusions.

One of most amazing insights that can be generated as a result of Figure 5.3 is to notice that the scaling function seems to be quite irregular. Indeed, if we look at the fine-scale of this function as illustrated in Figure 5.4, we can see that the scaling function for this very simple wavelet seems to be self-similar at different scales, i.e. it is a fractal. Indeed, the classical method for generating fractals is as the fixed point of an iterated function system on a space with a Hausdorff metric. See Barnsley (1988) for many more details on fractal geometry. Indeed the first equation of (5.5) together with (5.6) shows that  $\phi$  is indeed a fixed point of an iterated function system. The wavelet representation itself consists of an infinite sum of translates and dilates of a fundamental function  $\psi$ . It is, therefore, not surprising that there is a rather deep connection between wavelet analysis and fractals. Both wavelets and fractal geometry are at the leading edge of contemporary mathematics. We cannot hope to summarize both in one article let alone their connections at their theoretical roots. Nonetheless, there is an intriguing connection between the two.

It is perhaps best to summarize by reiterating the connections we have established. We began by addressing the immensely popular and fashionable topic of nonparametric function estimation. We showed that this can be cast as an optimization problem in functional analysis. The statistical solution to this optimization problem can be formulated by finding spanning bases for admissible classes of functions. Wavelets form a rather flexible base because of their doubly indexed nature. That is to say, where there is considerable fine-structure in the function to be estimated, the dilation structure can be made very fine for high resolution. But where there is smooth structure one or two simple dilations are sufficient. This localizing property has important implications for non-smooth function estimation and opens up a range of possibilities not available to the conventional convolution smoothers that dominate the current statistical literature.

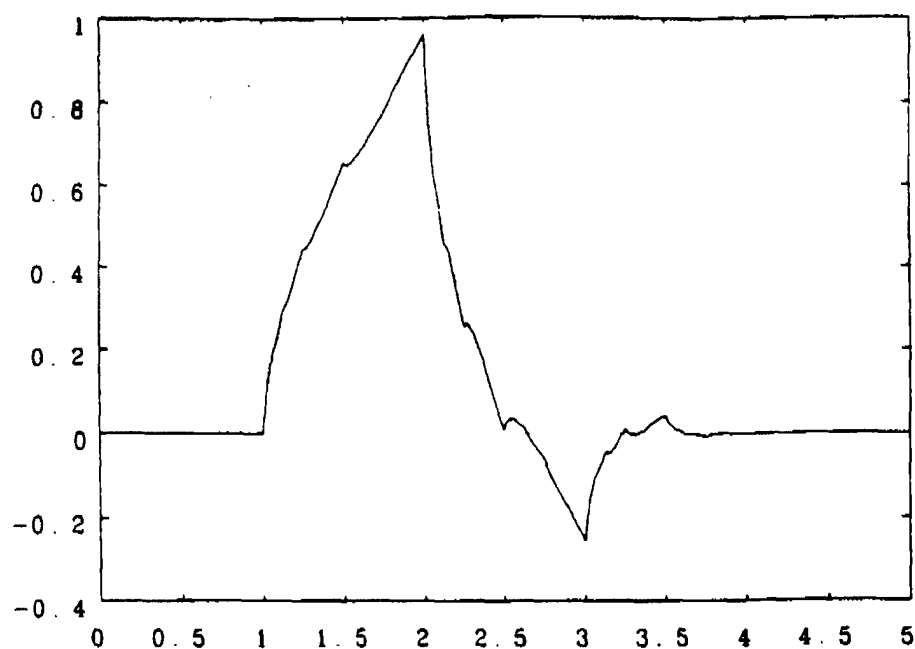


Figure 5.3a. Daubechies' Scaling Function using 4-term FIR filter.

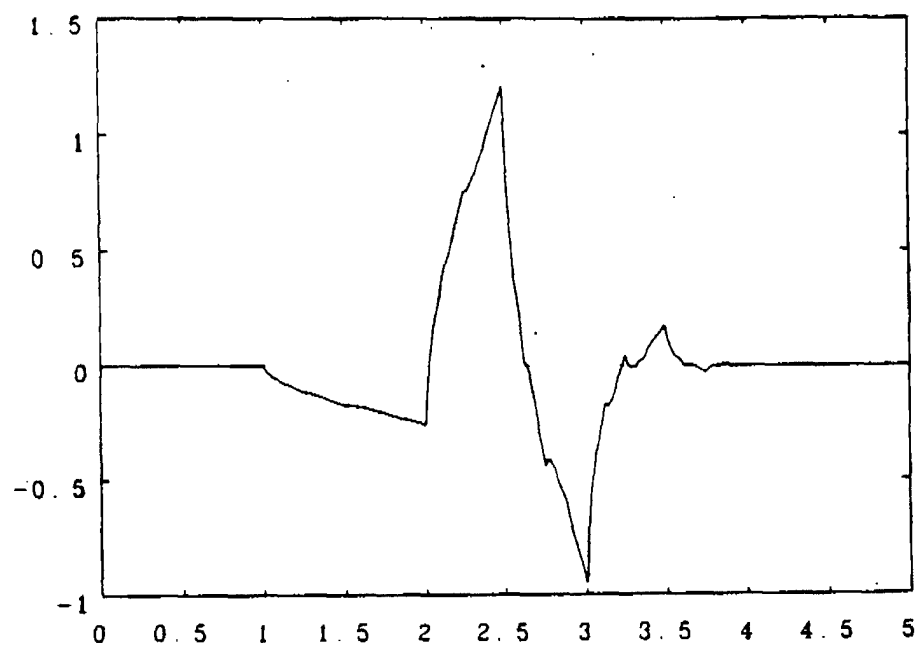


Figure 5.3b. Daubechies' Mother Wavelet using 4-term FIR filter.

Next in exploring wavelets in a somewhat deeper fashion, we have seen that they have properties as time-frequency generalizations of conventional harmonic analysis. In this context they form a methodology for analyzing nonstationary second-order time series. Indeed, the connection with signal processing is even deeper in that the signal processing methodologies known as sub-band coding and multiresolution analysis lead to a formulation of a constructive algorithm for both wavelet decomposition (and reconstruction) and for computing both scaling functions and mother wavelets. A theory which links nonparametric function estimation, functional analysis, nonstationary time series analysis, multiresolution signal processing, wavelet analysis and fractals cannot help but be intriguing. I hope this discussion will stimulate further interest. One final tidbit to entice further interest. We have seen in our discussion of multiresolution analysis that the  $V_j$  form a nested sequence of lower-resolution spaces (smoother spaces). Splines, in particular smoothing B-splines, form a method for low pass filtering; thus for constructing a multiresolution analysis. Thus splines can be used to construct wavelet bases. Moreover, certain classes of wavelets not only span  $L_2(\mathbb{R})$ , but also span Sobolev spaces of finite order. In particular, therefore, splines, which are optimizers over Sobolev spaces, may be written in terms of a wavelet basis. Thus to the list of rather deeply interconnected topics may be added the topic of splines.

**Acknowledgements.** My graduate students, Qiang Luo and Don Faxon, have been most helpful at providing graphic examples of wavelets at various stages in the development of this paper. I also benefited from conversations with Christian Houdré who, in particular, drew my attention to the implications of the paper of Duffin and Schaeffer.

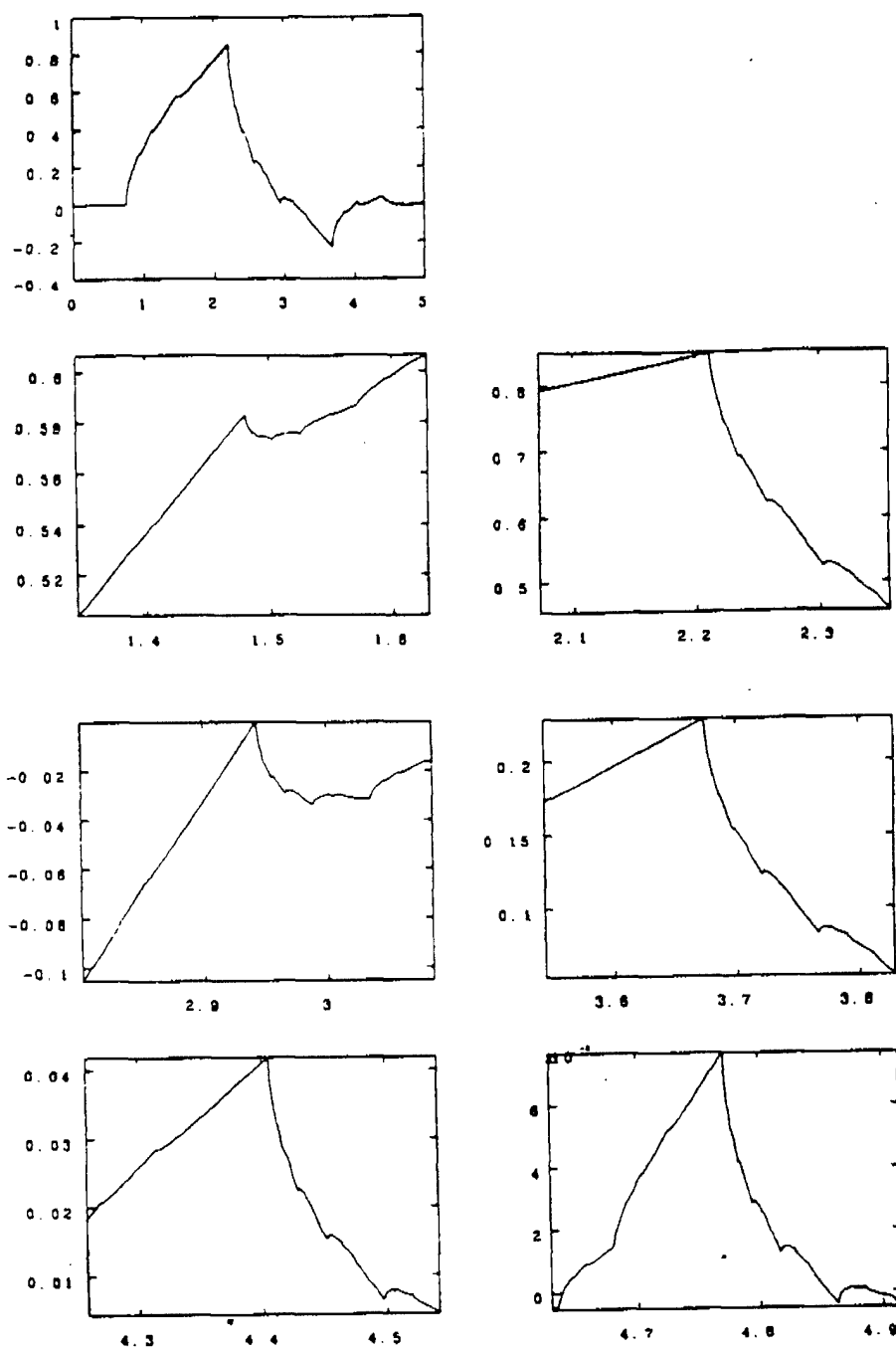


Figure 5.4 Daubechies' original 4-term FIR-based scaling function and six zoomed subgraphs showing fractal character of this scaling function. Several of the subgraphs have been inverted in order to illustrate the similarities.

## References.

- Barnsley, M. (1988), *Fractals Everywhere*, Academic Press: Boston
- Chui, C. K. (1992), *An Introduction to Wavelets*, Academic Press: Boston
- Daubechies, I. (1988), "Orthonormal bases of compactly supported wavelets," *Comm. on Pure and Appl. Math.*, 41, 909-996
- Daubechies, I., Grossmann, A. and Meyer, Y. (1986), "Painless nonorthogonal expansions," *J. Math. Phys.*, 27, 1271-1283
- Duffin, R. J. and Schaeffer, A. C. (1952), "A class of nonharmonic Fourier series," *Trans. Amer. Math. Soc.*, 72, 341-366
- Heil, C. and Walnut, D. (1989), "Continuous and discrete wavelet transforms," *SIAM Review*, 31, 628-666
- Le, H. T. and Wegman, E. J. (1991), "Generalized function estimation of underwater transient signals," *J. Acoust. Soc. America*, 89, 274-279, 1991
- Rioul, O. and Vetterli, M. (1991), "Wavelets and signal processing," *IEEE Sign. Proc. M.*, 8, 14-38
- Wegman, E. J. (1984), "Optimal nonparametric function estimation," *J. Statist. Planning and Infer.*, 9, 375-387.

# **A parallelized, simulation based algorithm for parameter estimation**

**Martin Lawera**

**James R. Thompson**

Rice University, Houston, Texas

## **Abstract**

The SIMEST algorithm for obtaining estimates of the parameters characterizing a stochastic process is implemented using a piecewise quadratic approximation to a goodness of fit statistic. The implementation is motivated in part by the rotatable experimental designs of Box and Hunter. Here, however, an "experiment" is simply a computer simulation, so the cost of the experiment is, essentially, trivial. Parallelized computation is used on a Levco transputer system choosing design points in a fashion so as to maximize the utilization of all transputers and the information obtained from the simulated data

## **1 Introduction.**

### **The motivation for and an overview of SIMEST**

Deep modeling of a stochastic phenomenon might properly begin with a basic understanding of the phenomenon expressed as a set of simple axioms at the micro level. In the best of all worlds, the next step would be to write the likelihood equation, giving the characterizing parameters as a (generally complex) function of the data. The third step then, is to find values of the parameters which approximately maximize the likelihood.

The first and third steps are much the easiest. It is the writing down of the likelihood which causes the trouble. Experience from biostatistics and other fields shows clearly the limitations of this traditional approach. Many real-life situations, can indeed be easily microaxiomitized. However, when we set out to translate those axioms into a likelihood function, we almost

always fail. Writing down the likelihood function requires us to delineate all possible pathways which could have produced the data set. If, as is usually the case, one state of affairs can be reached by very many different paths, we are very likely to get lost.

Traditionally the problem of going from the “forward” axioms to the “backward” likelihood equations is treated as intractable to be replaced by an “empirical” (e.g., linear or log-linear) model. Such “empiricism” has enabled economists to predict ten of the last three recessions.

SIMEST is a strategy which enables us to use the deep modeling approach and estimate the characterizing parameters, without the necessity of writing down the likelihood function. This is achieved by assuming a value for the parameters of the model and then generating simulated pseudo-data which are compared with the actual data. This comparison enables us to update our estimate for the parameters. This is, briefly, the idea of SIMEST.

In more detail, the SIMEST approach consists of three elements. First, based on the principles of the process we are modeling, we develop a simulation. We then use this simulation to produce “pseudo-data”, i.e., simulated values which can be compared to the real data.

Secondly, we use some goodness-of-fit function to quantify the conformity between the simulated and the real data. Pearson’s  $\chi^2$  statistic:

$$\chi^2(\Theta) = \sum_{i=1}^k \frac{(\hat{p}_i(\Theta) - p_i)^2}{\hat{p}_i(\Theta)}$$

is a good candidate here. Other candidates and the criteria for choosing among them have been discussed in detail by Thompson, Brown and Atkinson [3] and Thompson [4].

Thirdly, we perform the simulation for various sets of parameters and use an optimization procedure to find the parameter set which produces results the closest to the real data. Those values are taken to be the SIMEST estimates.

## 2 Example:

### Cancer progression model

The difficulties resulting from the classical approach to estimation are well exemplified with the cancer progression model of Bartoszynski, Brown and

Thompson [1].

In this case, we are studying a population of patients who had been diagnosed with cancer (the primary tumor) at a certain time  $tD$ . The patients had undergone surgery which removed the primary. Nevertheless, after a period of time they were diagnosed with cancer again (the secondary tumor). We want to know whether the secondary had originated from a metastasis of the primary, which was not noticeable at the time of surgery, or whether it had independently been produced by the systemic mechanism.

Based on clinical experience, we assume that the metastatic and the systemic processes obey the following simple axioms:

1. For each patient, each tumor originates from a single cell, and grows exponentially at rate  $\alpha$ .
2. The probability of systemic occurrence of a tumor in  $(t, t + \Delta t)$  equals  $\lambda * \Delta t + o(\Delta t)$ , independent of the prior history of the patient.
3. The probability that a tumor not previously detected, will be detected and removed in  $(t, t + \Delta t)$  is  $b \times Y(t) + o(\Delta t)$ , where  $Y(t)$  is the size of the tumor at  $t$ .
4. Until the removal of the primary, the probability of a metastasis in  $(t, t + \Delta t)$  is  $\beta \times Y(t) + o(\Delta t)$ .

Since both processes as described by those four axioms are modified Poisson processes with intensities growing exponentially in time, it is straightforward to find the distribution functions of the random variables involved. Using

$tD$	for the discovery time of the primary
$tM_i$	for the occurrence time of the $i$ th metastatic secondary tumor
$tS_i$	for the occurrence time of the $i$ th secondary systemic tumor
$td$	for the discovery time of any secondary tumor



we have:

$$\begin{aligned} F_D(tD) &= 1 - \exp\left(-\int_0^{tD} b e^{\alpha\tau} d\tau\right) \\ &= 1 - \exp\left(-\frac{b}{\alpha} e^{\alpha tD}\right) \end{aligned}$$

$$\begin{aligned} F_M(tM_i) &= 1 - \exp\left(-\int_{tM_{i-1}}^{tM_i} a e^{\alpha\tau} d\tau\right) \\ &= 1 - \exp\left[\frac{a}{\alpha}(e^{\alpha tM_{i-1}} - e^{\alpha tM_i})\right] \end{aligned}$$

$$F_S(tS_i) = 1 - e^{-\lambda tS_i}$$

$$\begin{aligned} F_d(td) &= 1 - \exp\left(-\int_0^{td} b e^{\alpha\tau} d\tau\right) \\ &= 1 - \exp\left(-\frac{b}{\alpha} e^{\alpha td}\right) \end{aligned}$$

Continuing from this point on using the traditional maximum likelihood methods has proved quite difficult. As showed by Bartoszynski, Brown and Thompson [1], the likelihood equations are solvable in closed form only after dropping the discovery times for the secondary tumors, i.e. after assuming that

$$tdS_i = tS_i, \quad tdM_i = tM_i.$$

And even then, the computations involved solving a three dimensional quadrature: a massive task, even for a modern computer.

On the other hand, obtaining a SIMEST solution to this problem is straightforward. Knowing that random variates from the above distributions can easily be generated from  $t = F^{-1}(u)$ , where  $u$  is a  $U(0,1)$  random variate, we use the following flowchart to simulate the discovery times for secondary tumors:

1. Generate  $tD$
2. Generate the sequence  $\{tM_i, 1 \leq i \leq n\}$ , such that  $tM_{n+1} > tD$
3. Generate the corresponding sequence  $\{tdM_i, 1 \leq i \leq n\}$
4. If  $tdM_i, 1 \leq i \leq n$ , is less than  $tD$ , set  $tdM_i = \infty$
5. Generate the sequence  $\{tS_i, 1 \leq i \leq m\}$ , such that  $tS_{m+1} > 5 \times tD$
6. Generate the corresponding sequence  $\{tdS_i, 1 \leq i \leq m\}$
7. If  $tdS_i, 1 \leq i \leq m$ , is less than  $tD$ , set  $tdS_i = \infty$
8. Output  $\min(\{tdM_i, 1 \leq i \leq n\}, \{tdS_i, 1 \leq i \leq m\})$

Sacrificing some of its simplicity, we may make this flowchart more efficient by halting the generation of  $tS_i$  at  $i = k$  if, for some  $j < k$ :

$$tdS_j > tD \quad \text{and} \quad tS_k \geq tdS_j$$

Since, as we have found out, the whole process is dominated by the systemic tumors, this change significantly reduces the running time.

On the next stage of the SIMEST procedure, we repeat the simulation a larger number of times for various parameter sets. The outputs from these simulations form "pseudo-data" sets and are compared to the actual data. The similarity between the real and the simulated data is quantified by Pearson's  $\chi^2$  statistic and used by an optimization routine to search for the set of parameters maximizing the goodness of fit to the observations.

### 3 An optimization algorithm for noisy functions

It is clear from the above overview of SIMEST that an efficient and robust optimization procedure is crucial for a successful implementation of this method. Unfortunately, most of available optimization routines are ill-suited for the task. The goodness of fit statistic that we are trying to minimize is based on simulated data and therefore is contaminated with noise. Obviously, presence of the noise may be catastrophic for all Newton-type approaches.

Direct search methods, such as Nelder-Mead algorithm or STEPIT are better candidates for application in SIMEST, and in fact, have been successfully used. These methods, although not designed specifically for noisy functions, are usually robust enough to handle the task, although they do occasionally converge prematurely, or veer in a completely wrong direction. Furthermore, when using those procedures, we try to keep the noise down by averaging a very large number of simulations which greatly increases the running time.

Such difficulties prompted us to develop an optimization routine for noisy functions. It attempts to combat noise by increasing the number of simulations at the points where it is necessary and to avoid a direct search by finding a minimum through a local quadratic approximation to the goodness-of-fit function. Unlike other procedures, however, ours calculates the derivatives based on regression and not on finite differences. The design matrix for regression is one of the Box and Hunter "cube-and-star-points" rotatable design.

### General Description:

The algorithm starts from an initial guess, provided by the user. Around this point, we set up a rotatable design, consisting of  $2^n + 2 \times n + n_0$  points, where  $n_0$  is the number of replicates of the center (see Figure 1). For example, in the two dimensional case, with  $n_0 = 2$ , the design points have the following coordinates:

$(0, 0), (0, 0)$  "center"

$(1, 1), (-1, -1), (1, -1), (-1, 1)$  "cube"

$(1.682, 0), (-1.682, 0), (0, 1.682), (0, -1.682)$  "star"

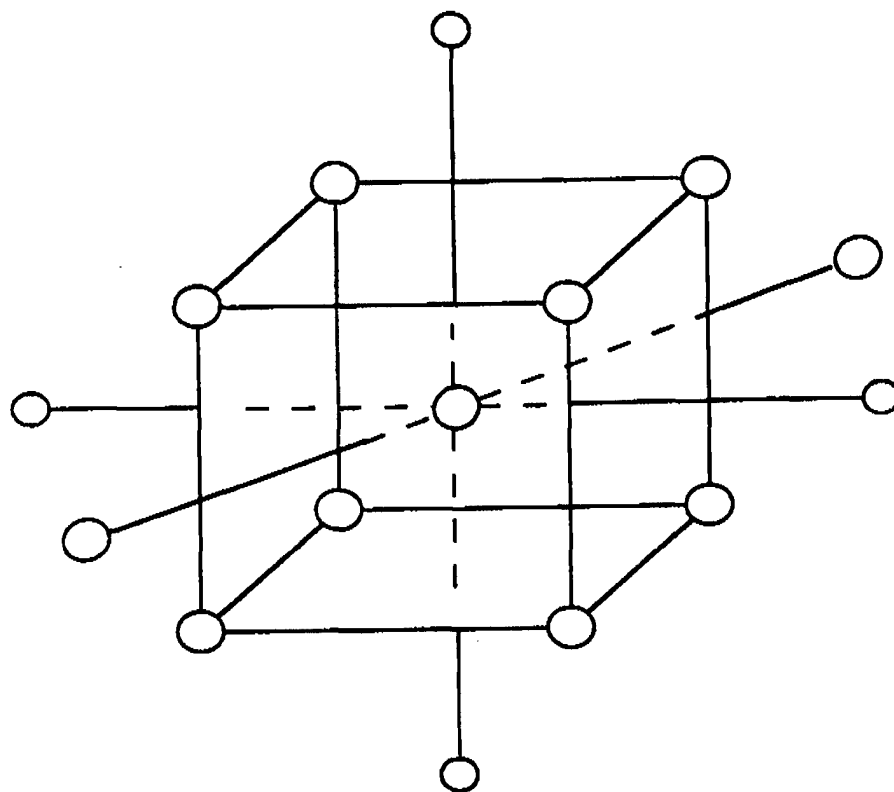
in the units of the design.

Next, we calculate the "absolute" coordinates of the design points and proceed to evaluate them.

Evaluation of each point is itself a staged process in an effort to eliminate the noise. First seven evaluations (a detail due to the parallel implementation) are taken and their mean and the variance are calculated. If the variance of the mean is above a prespecified level, another seven values are found and

Figure 1

Box-Hunter "cube and star points" rotatable  
design in three dimensions.



the process continues until the variance of the mean is low enough, and then the mean is taken as the value of the objective function at this point.

When all the design points have been evaluated, we use least squares to fit a quadratic polynomial to them:

$$J_1(\Theta) = \beta_0 + \sum_{i=1}^p \beta_i \Theta_i + \sum_{i=1}^p \sum_{j=1}^p \beta_{ij} \Theta_i \Theta_j$$

and then transform the polynomial to canonical form A:

$$J_2(\Theta) = \beta_0 + \sum_{i=1}^p \beta_i \Theta_i + \sum_{i=1}^p \beta_{ii} \Theta_i^2$$

What happens next depends on how good the quadratic fit is, as measured by the  $r^2$  statistic. Suppose the fit is good ( $r^2 > 0.9$ ). Then, based on the quadratic approximation, we calculate three quantities:

1. the minimum
2. the new rescaling matrix
3. the new rotation matrix.

They will be used in the next iteration, unless a better alternative is found further on during the present step. The minimum is calculated in the normal way, by setting the derivatives equal to zero. However, if it turns out to lie far outside of the trust region, we replace it with a point within this region, obtained by setting the smallest in the absolute value coefficients in the form A to zero. Next, the minimum is evaluated using the same technique as for the design points.

The rotation matrix is calculated to ensure that the axes of the new design will coincide with the axes of the fitted hyper-ellipsoid.

The new rescaling matrix rescales each variable by the proportion of the total variation of the objective function contributed by this variable. It ensures that in the units of the new design, the fitted hyper-ellipsoid will approximately be a hyper-sphere.

If the fit is unsatisfactory ( $r^2 < 0.9$ ), we proceed directly to the next stage (still within the same iteration) which is to shrink or to expand the design without changing its center.

Such a uniform rescaling of the design is an attempt to find a better quadratic fit. The present design can be suboptimal for two reasons. First, it can be too large, so that the objective function is too variable to be approximated by a quadratic polynomial. This case will be manifested by low  $r^2$  and a high range of the function values and a high Error Sum of Squares from the regression, in comparison to the level of noise. Then, we divide each element of the rescaling matrix by 2 and go back to the evaluation stage.

On the other hand, if the range of values obtained from the design is small relative to the noise and so is the Error Sum of Squares from the regression; and if, at the same time the fit is still bad, we would conclude that the design is not spread out enough. In this case, there is no quadratic or even linear effect in the objective function so that the Explained Sum of Squares is very low. There is some variation though due to the noise and it brings the Error Sum of Squares up. Notably, the latter kind of variation cannot be eliminated since we can only control variation "within" a design point, and not between points. Hence, the only solution is to expand the design, so as to give the variability of the function itself a chance to show.

Consequently, if at the beginning of an iteration, we get an unsatisfactory fit, we can identify the reason, and act appropriately by either shrinking or expanding the design. If the first fit is good, we could use it right away and make it the center of a new design, thus beginning another iteration. It seems better though to avoid hopping around, and to try both to shrink and to expand the initial design hoping for an ever better fit and a more substantial decrease in the minimum value.

The stopping criteria for the rescaling stage are natural: shrinking must stop if we arrive at a design which cannot be fit well, apparently because it is too small. Similarly, the expansion stops when the design becomes too large to be approximated well with a quadratic. In the latter case, an additional restriction is needed since if the objective function actually is quadratic, no design will be too large. For this reason, we also impose an upper bound on the number of expansions.

At the last stage of each iteration we choose the center of the design to be evaluated in the next iteration. If the smallest value obtained during the present iteration is significantly smaller than the value at the present minimum, i.e if the expected gain from moving to that point exceeds the level of noise, we center the new design there. Otherwise, we do not move, but start the new iteration like the previous one, except for the upper bound

on the level of noise, which is reduced four times.

The algorithm stops if through subsequent reductions the upper bound on the noise has been reduced to a certain minimum, and despite that, no move is possible.

## Flowcharts

We will present now a series of flowcharts which correspond to five levels of the algorithm, arranged in the descending order of generality. Hence, the upper levels will consist to a large extent in a repeated executions of the levels below them.

The following objects will be referred to:

- $\Theta_0$             vector of coordinates of the current minimum,  
                  i.e. the center of the currently fitted design
- $D$                  $n \times (2^n + 2n + n_0)$  Box-Hunter design matrix, where  $n$  is the  
                  dimensionality of the problem and  $n_0$  is the number of replicates  
                  taken at the center.
- $R$                  $n \times n$  diagonal matrix used to transform the units of the design  
                  into the units of the "absolute" coordinate system
- $T$                  $n \times n$  matrix which rotates the axes of the design  
                  to coincide with the "absolute" axes

**Note:** The design points as given by matrix  $D$  have "absolute" coordinates given by:

$$T \times R \times D + \Theta_0$$

- $S(\cdot)$             the objective function
- EC                presepecified by the user upper limit on the level of noise
- CONV            user-specified constant in the convergence criterion at level 2
- $X_R$              $(2^n + 2n + n_0) \times [1 + n + n(n + 1)/2]$  matrix of regression points

X can be written as:

$$[\mathbf{1}, (R \times D)^T, rd_{11}, \dots, rd_{nn}]$$

where  $\mathbf{1}$  is a column of 1's and

$$rd_{ij} \quad 1 \leq i \leq n, \quad i \leq j \leq n$$

is a column vector obtained by elementwise multiplication of the  $i$ th and the  $j$ th columns of  $(R \times D)^T$

### Level 1

**Input (initial guess):**  $\Theta_0, R_0, T_0$

1. Perform the level 2 optimization starting from the initial guess.  
Output:  $\Theta_1, R_1, T_1$
2. Perform the level 2 optimization ten times, starting each time from the results obtained in (1)
3. Find  $S_{min}(\Theta_{min})$ : the best of results obtained in (2)

**Ouput:**  $\Theta_{min}, S_{min}(\Theta_{min})$

### Level 2

**Input:**  $\Theta_0, R_0, T_0, EC_0$

1.  $EC \leftarrow EC_0$
2. Perform the level 3 optimization using the input values.  
Output:  $\Theta_1, R_1, T_1, S(\Theta_1)$
3. Calculate the distance between  $\Theta_0$  and  $\Theta_1$

$$\Delta\Theta = \sqrt{\|\Theta_1 - \Theta_0\|^2}$$



4. Calculate the gain from (2):

$$\Delta S = S(\Theta_0) - S(\Theta_1)$$

5. If

$$\Delta\Theta > 0 \quad \text{and} \quad \Delta S > 1.5 \times \sqrt{EC}$$

then

$$\Theta_0 \leftarrow \Theta_1, \quad R_0 \leftarrow R_1, \quad T_0 \leftarrow T_1, \quad \text{and} \quad \text{Goto (1)}$$

6. Else if

$$EC > CONV$$

then

$$EC \leftarrow EC/4, \quad \text{and} \quad \text{Goto (2)}$$

7. Else exit to level 1.

**Output:**  $\Theta_1, R_1, T_1, S(\Theta_1)$

### Level 3

**Input:**  $\Theta_0, R_0, T_0, EC_0, S(\Theta_0)$

1. Perform level 4. Output:  $R_{min}$ .

2. Set

$$Y \leftarrow NULL, \quad X \leftarrow NULL, \quad i \leftarrow 0, \quad \Theta_1 \leftarrow \Theta_0$$

3. Set  $R_{Cur} \leftarrow R_{min}$

4. Increment  $i$  by one

5. Evaluate

$$S(\Theta)^* = S(T_0 \times R_{Cur} \times D + \Theta_1)$$

6. Calculate  $X_{R_{Cur}}$

7. Set

$$X \leftarrow \begin{bmatrix} X \\ X_{R_{Cur}} \end{bmatrix}$$

8. Set

$$Y \leftarrow \begin{bmatrix} Y \\ S(\Theta)^* \end{bmatrix}$$

9. Regress  $Y$  on  $X$ .

Obtain: vector of regression coefficients  $\hat{\beta}$  and the  $r^2$  statistic.

10. Perform the level 5 optimization.

Output:  $\Theta^*$ ,  $R^*$ ,  $T^*$ ,  $S(\Theta)^*$

11. Calculate the gain from (5):  $\Delta S = S(\Theta_1) - S(\Theta)^*$

12. If

$$r^2 > 0.9, \quad \Delta S > 1.5 \times \sqrt{EC}, \quad i < 20$$

then

$$\Theta_1 \leftarrow \Theta^*, \quad R_1 \leftarrow R^*, \quad T_1 \leftarrow T^*$$

$$R_{Cur} \leftarrow 2 \times R_{Cur}, \quad \text{and} \quad \text{Goto (4)}$$

13. Else exit to level 2

**Output:**  $\Theta_1, R_1, T_1, S(\Theta_1)$

**Level 4**

**Input:**  $\Theta_0, R_0, T_0, EC_0, S(\Theta_0)$

1. Evaluate

$$S(\Theta_0)^* = S(T_0 \times R_0 \times D + \Theta_0)$$

2. Calculate  $X_{R_0}$

3. Regress  $S(\Theta_0)^*$  on  $X_{R_0}$

Obtain: the  $r^2$  statistic and the Error Sum of Squares (ESS)

4. If

$$r^2 < 0.9, \text{ and } (ESS < 2 \times EC_0, \text{ or } \text{Max}(S(\Theta_0)) - \text{Min}(S(\Theta_0)) < 1.5 \times \sqrt{EC})$$

then

(a) Set  $R_0 \leftarrow 2 \times R_0$

(b) Repeat (1) - (4) until

$$r^2 > 0.9, \text{ or } (ESS > 2 \times EC_0, \text{ and } \text{Max}(S(\Theta_0)) - \text{Min}(S(\Theta_0)) < 1.5 \times \sqrt{EC})$$

(c) Exit to level 3

5. Else

(a) Set  $R_0 \leftarrow 0.5 \times R_0$

(b) Repeat (1) - (4) until

$$r^2 < 0.9, \text{ and } (ESS < 2 \times EC_0, \text{ or } \text{Max}(S(\Theta_0)) - \text{Min}(S(\Theta_0)) < 1.5 \times \sqrt{EC})$$

Set  $R_0 \leftarrow 2 \times R_0$

(c) Exit to level 3

**Output:**  $R_0$

Level 5

**Input:**  $\hat{\beta}$ , quadratic fit to the objective function

**Note:** This part is based on Box and Draper[2]

1. Calculate vector  $b$  and the matrix  $B$  such that the quadratic fit has the form:

$$\hat{y} = b_0 + X^T \times b + X^T \times B \times X$$

2. Find matrices  $M$  and  $\Lambda$  such that  $M^T \times B \times M = \Lambda$
3. Calculate the minimum  $\Theta \leftarrow -1/2B^{-1} \times b$

4. If  $\sqrt{\|\Theta\|^2} > 1$  then
  - (a) Set  $Min(|\Theta_1|, \dots, |\Theta_n|) \leftarrow 0$
  - (b) Repeat (a) until  $\sqrt{\|\Theta\|^2} \leq 1$
5. Calculate the rescaling matrix  $R_0 \leftarrow Diag(|\lambda_i|^{-1/2})$
6. Set  $T_0 \leftarrow M$

**Output:**  $\Theta_0, R_0, T_0$

### Evaluation

**Input:**  $\Theta, R, T, EC$

1. Set  $i \leftarrow 0$
2. Evaluate  $S$  10 times at  $\Theta_0 = T \times R \times D + \Theta$
3. Increment  $i$  by 7
4. Calculate the sample mean  $\bar{S}$  and the sample variance  $V$  of all  $i$  evaluations
5. If  $V/EC > C_{i-1}$ , where  $C_{i-1}$  is the 95th percentile of the  $\chi^2_{i-1}$  distribution, then Goto (2)
6. Else exit

**Output:**  $\bar{S}$

## 4 Parallel implementation

One of the great advantages of the Box-Hunter rotatable design is its suitability for implementation on a parallel machine. For any change to this design, i.e. rescaling it or moving its center affects all points in the same way, so that what happens at one point is not dependent on what is happening at that time at another one.

Notably, as it is applied in our algorithm, the rotatable design can be parallelized in a number of ways. The choice among them depends mostly on the number of nodes that are available and on the dimensionality of the problem.

Referring back to the cancer progression model, the problem we are solving there is four dimensional. This translates into  $16 + 8 + 2 = 26$  design points. If we are using so many nodes that the design points can be approximately evenly split among them, we can assign a point or a group of points to each node. All the evaluation is then done independently on each node, then the nodes work together to fit a quadratic, find a minimum and decide upon rescaling the design. Then, the nodes work separately again.

If the number of available nodes is several times higher than the number of design points, another strategy is called for. We would assign nodes to individual points in several concentric designs. If the number of such layers is large enough, we can substitute this strategy for rescaling the design. We would then simply compare the minima obtained in each layer and pick the best one.

Furthermore, if the number of available nodes does not allow for an easy division of work among them, we can come up with a more elaborate sharing scheme, so that nodes which finished their jobs sooner could help out those still working. This, however, requires designating one node as the "group leader", picking information from all the nodes and reassigning their tasks as the need arises. Consequently, the programming aspect becomes much more complicated, and some of the running time is wasted on the additional "hand-shaking".

Additional consideration comes from the proportion of time necessary to do the evaluations to the time necessary for calculating the minimum, rescaling the design, etc. As our experience indicates, in the case of a simulation-based goodness-of-fit function, doing function evaluations takes much more time than all the remaining tasks combined. Taking into account that we were

working on just seven nodes, we decided that designating one of them as a "group leader" and thus taking it off evaluation, would be counterproductive.

We were still left with the possibility of assigning 4 design points to each of 6 six nodes and just 2 points (perhaps those in the center) to the remaining node. This would work reasonably well for a four dimensional problem, but not so well for two or three dimensional ones (having 10 and 16 design points respectively). Hence, for the sake of versatility, we chose yet another approach.

We decided to place parallelism directly on the evaluation level, and have all nodes work jointly to evaluate each design point. This is the reason for taking seven evaluations of each point at a time, which we mentioned above. This setup does not require any changes for handling problems of various dimensionalities and provides the best use of the resources we used. Naturally, under different circumstances one of the possibilities described above may be more advantageous.

## 5 Numerical examples

To test our algorithm, we applied it to three optimization problems. The starting values, the true minima and the estimates we obtained are given below. In addition, Figures 2 and 3 contain scatter plots showing intermediary steps of the algorithm between the starting point and the final estimate.

Firstly, we used a simple quadratic function of three variables with additive Gaussian noise with the zero mean and the variance of 5.

$$J_1(\Theta) = \Theta_1^2 + \Theta_2^2 + \Theta_3^2 + \epsilon$$

The starting point was:

5.000000	3.000000	-2.000000
----------	----------	-----------

The initial estimate:

-0.000000	-0.000000	0.000000	value = 0.170180
-----------	-----------	----------	------------------

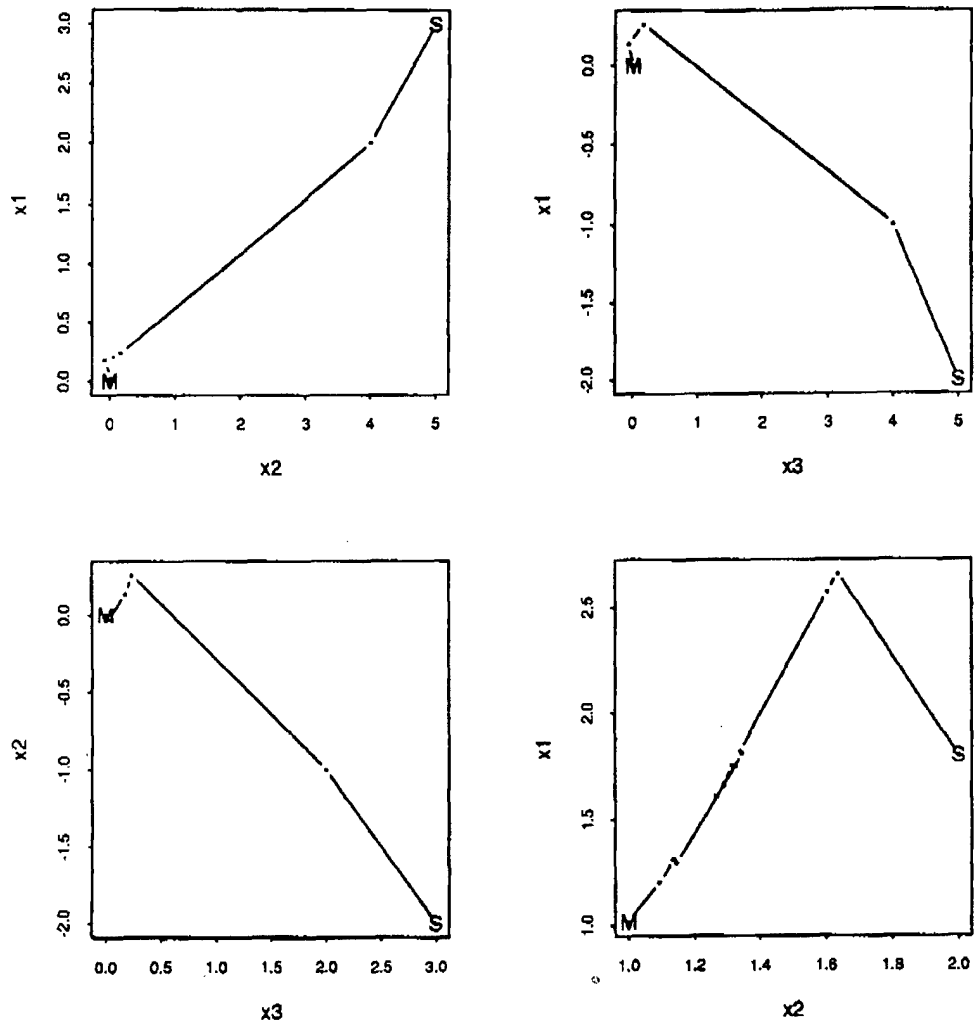
The final estimate:

-0.000000	-0.000000	0.000000	value = -0.062558
-----------	-----------	----------	-------------------

The true minimum:

0.000000	0.000000	0.000000	value = 0.000000
----------	----------	----------	------------------

Figure 2.  
Optimization Paths for the Test Functions

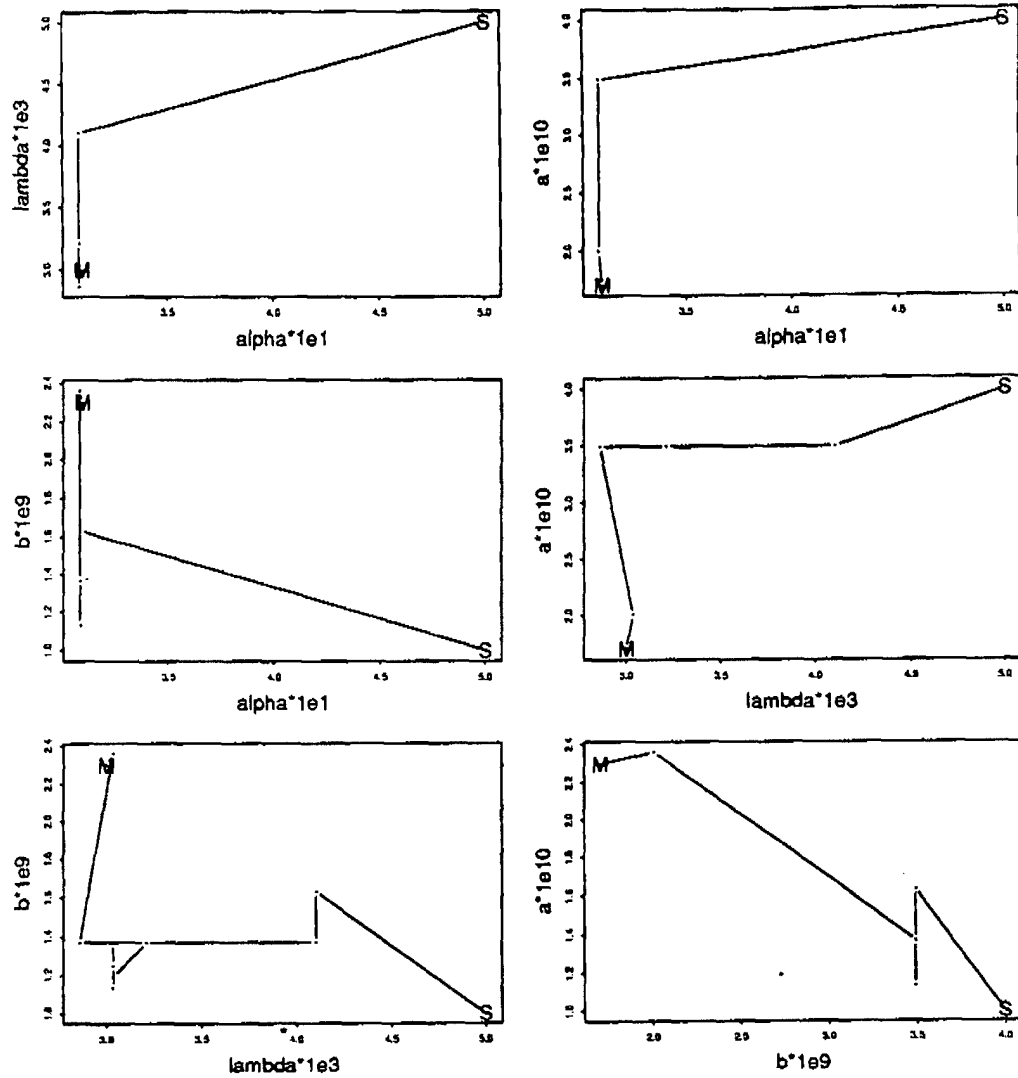


First three graphs show the optimization paths for the quadratic function,  
the fourth one, for the Rosenbrock function.

S is the starting point, M is the true minimum

Figure 3.

# Optimization Paths for the Cancer Progression Model



S is the starting point, M is the minimum



Secondly, we used the Rosenbrock function, with additive standard Gaussian noise (i.e.,  $N(0, 1)$ ).

$$J_2(\Theta) = 100 * (\Theta_1^2 - \Theta_2)^2 + (1 - \Theta_1)^2 + 1 + \epsilon$$

The starting point was:

2.000000                      1.800000

The initial estimate:

1.095556                      1.203521                      value = 1.025287

The final estimate:

0.999584                      1.017149                      value = 0.992749

The true minimum:

1.000000                      1.000000                      value = 1.000000

Finally, we simulated 150 "pseudo-patients" from the cancer progression model outlined above and then tried to recover the original parameters:

The starting point was:

$\alpha = 5.0e - 1$      $\lambda = 5.0e - 3$      $a = 4.0e - 10$      $b = 1.0e - 9$

The initial estimate:

$\alpha = 3.1e - 1$      $\lambda = 2.8e - 3$      $a = 3.5e - 10$      $b = 1.4e - 9$     value = 6.9336

The final estimate:

$\alpha = 3.1e - 1$      $\lambda = 3.2e - 3$      $a = 2.0e - 10$      $b = 2.3e - 9$     value = 6.6280

The true minimum:

$\alpha = 3.1e - 1$      $\lambda = 3.0e - 3$      $a = 1.7e - 10$      $b = 2.3e - 9$

## 6 Conclusions

SIMEST is an estimation strategy which effectively bypasses the difficulties involved in writing down the likelihood function and with solving the likelihood equations. For the full success, SIMEST requires an optimization procedure which would retain its reliability in the presence of noise but which could also avoid unreasonably long running times.

The procedure presented here has those advantages. By increasing the number of evaluations when necessary, it ensures a low level of noise. By

using regression instead of finite differences, it gives a better picture of the function's variation. Finally, it is suited for easy parallelization, as it is based on rotatable designs.

## 7 References

1. Bartoszynski, Robert, Brown, B.W. and Thompson, J.R. (1982) "Metastatic and systemic factors in neoplastic progression." *Probability Models and Cancer*. Eds. L. LeCam J. Neyman. New York: North Holland. 253-264
2. Box, G.E.P., Draper, N.R. *Empirical Model Building and Response Surfaces*. (1987). John Wiley and Sons. 304-380
3. Thompson, J.R., Atkinson, E.N. and Brown, B.W. (1987). "SIMEST: an algorithm for simulation based estimation of parameters characterizing a stochastic process." *Cancer Modelling*. Eds Thompson, J. and Brown B. New York: Marcel Dekker. 387-415.
4. Thompson, J.R. *Empirical Model Building*. (1989). John Wiley and Sons. 114-131

## 8 Acknowledgments

This work was supported in part by the Army Research Office, under DAAL-03-91-G-0210, and DAAL-03-88-G-0074

## Simulation Based Estimation for Birth and Death Processes

*Katherine B. Ensor*  
Department of Statistics  
Rice University  
Houston, TX 77251-1892

*Eileen Bridges*  
Jesse H. Jones Graduate School of Administration  
Rice University

*Martin Lawera*  
Department of Statistics  
Rice University

### ABSTRACT

We examine the applicability of simulation based estimation in complicated stochastic processes for which likelihood estimates are not a viable option. Implementation of this procedure is considered for homogeneous Poisson processes and birth and death processes with consideration given to the use of parallel computing. The statistical properties of the estimator are investigated with strong consistency demonstrated for the *i.i.d.* case. A new method of estimating the variance of the estimators is also suggested.

### 1. INTRODUCTION

Often it is the case that stochastic models are easily axiomatized from the properties of the physical process under study but it is extremely difficult to obtain the likelihood equation from these well understood axioms. Thus the scientist tends to use a deterministic approximation to the truth which can be solved more easily. If a stochastic component is still deemed necessary in the modeling attempts additive error models are usually considered. Simulation based estimation or SIMEST offers an alternative to the simplifying assumption of deterministic modeling.

The simulation based estimation method outlined here was originally motivated by the work of Thompson, Brown and Atkinson (1987) in modeling cancer progression. Models based on simple biological axioms at the micro level lead to likelihood functions of extreme complexity at the macro level at which clinical data is available. From the axioms of the physical process various probabilities of interest can be obtained which would then enable one to establish the likelihood equations for estimation purposes. However, these probabilities are extremely complex leading to optimization difficulties with the likelihood function. The premise of their work was to bypass the tedious development and optimization of the likelihood equations by estimating the model parameters and desired quantities directly from the model assumptions.

One class of problems which prove to be difficult or impossible to solve in closed form are birth and death models representing processes occurring commonly in epidemiology, sociology, and marketing (Eliashberg and Chatterjee (1986)). For example, in epidemiology, Bartlett's (1960) stochastic model of a measles epidemic includes a "birth" term

---

This research was supported in part by the Army Research Office under grants DAAL 03-91-G-0210 and DAAL 03-88-G-0074.

This paper was presented at the Thirty-Seventh Conference in this series.

representing increase in the number of infective persons, which is proportional to both the number of infective persons and the number of susceptible persons in the population, and a "death" term representing decrease in the number of infective persons due to either recovery or death. In sociology, the survival of a social group (such as a political party) is described as a stochastic process by Bartholomew (1982). The population is divided into "susceptibles" and "spreaders." The number of spreaders increases in proportion to the number of spreaders and susceptibles, and decreases (possibly only temporarily) when a person ceases to be an active spreader. The death term is the key difference between this model and that of the measles epidemic: in the sociological model, "death" may be temporary.

An example in marketing is offered by Bridges, Ensor, and Thompson (1992), who suggest that the number of products competing in a particular product category may be modeled as a birth-and-death process. Here, "births" are product entries and "deaths" are withdrawals from the marketplace. Tapiero (1975) develops a birth-and-death process model for sales as a function of advertising, in which "births," or increases in sales, are proportional to advertising expenditures and the number of remaining potential customers, and "deaths" occur when sales decrease due to customer "forgetting" of the brand. A closed form solution may be obtained only if a very simple functional form is assumed for advertising expenditures.

The SIMEST procedure provides an alternative to the complicated problem of establishing the functional form of the probabilities; namely, from the axioms defining the stochastic process  $m$  realizations of the process are generated at a particular  $\theta \in \Theta$ , where  $\Theta$  denotes the parameter space. For large  $m$ , the average of the simulated realizations will approximate the true average for the process. The simulated realizations are then compared to the observed data by a function measuring the disparity,  $S_n(\theta)$ . The estimator of  $\theta$  is the value  $\hat{\theta} \in \Theta$  for which  $S_n(\theta)$  is minimized.

In the original motivating work by Thompson *et. al.* (1987),  $\{N(t)\}$  represents the time until onset of a secondary tumor in women presented with breast cancer. From their database of 116 women presented with primary breast cancer they were able through the use of SIMEST to reliably estimate the model parameters thereby obtaining information on important questions such as the probability of metastasis prior to detection of the primary tumor. For a complete exposition of this problem see Thompson and Tapia (1990, Chapter 8). The SIMEST method of estimation is also applicable if one observes a single realization of a stochastic process, e.g. a birth and death process. The mean path of the process for a particular set of parameter values is simulated from the axioms defining the process and again, the concordance between the simulated mean path and the observed series indicates the viability of the current parameter values.

SIMEST has been used to successfully estimate parameters in both of the above mentioned scenarios (Bridges, Ensor and Thompson (1992)). It is the purpose of this paper to explore the statistical properties of the SIMEST estimator such as unbiasedness, consistency, and methods of variance estimation as well as the practical issues in implementing this procedure.

## 2. SIMEST FOR RANDOM SAMPLES

Consider first a random sample  $t_1, \dots, t_n$  of size  $n$  from the stochastic process  $\{W(s), s \geq 0\}$  which represents the waiting time until the  $s^{\text{th}}$  event. In other words, we are given  $n$  independent observations of the time until a particular event of the process occurs. To obtain parameter estimates it is necessary to simulate  $m$  observations from this process, or

$m$  simulated times until occurrence of the event of interest.

### 2.1 Criterion Function

How should the simulated series for a given set of parameters be compared to the observed data? Dividing the time axis into  $k$  bins, let  $\hat{p}_1, \dots, \hat{p}_k$  denote the proportion of the  $n$  observations falling into each bin. The number of observations falling into each bin follows a multinomial distribution with the probability of any given observation resulting in bin  $j$  given by  $p_j$  for  $j = 1, \dots, k$  and  $\sum_{j=1}^k p_j = 1$ . The value of  $p_j$  is determined by the defining stochastic process  $\{N(t)\}$  which depends on the parameter  $\theta$ . Whenever it is necessary to emphasize the dependence on  $\theta$  we will denote the true parameter values by  $p_j(\theta)$ , for  $j = 1, \dots, k$ . Now, let  $\bar{p}_1(\theta), \dots, \bar{p}_k(\theta)$  denote the proportion of the  $m$  simulated data points falling into the  $k$  bins. In other words, as an estimate of  $p_j(\theta)$  for a given set of parameter values we use  $\bar{p}_j(\theta)$ . The Pearson goodness-of-fit statistic is given by

$$S_n(\theta) = \sum_{j=1}^k \frac{(\bar{p}_j(\theta) - \hat{p}_j)^2}{\bar{p}_j(\theta)} \quad (2.1)$$

where  $\bar{p}_j(\theta)$  replaces  $p_j(\theta)$  for  $j = 1, \dots, k$ . The estimator of  $\theta$  is the value  $\hat{\theta} \in \Theta$  which minimizes  $S_n(\theta)$ .

Whenever one divides the data into bins for comparative purposes consideration must be given to the optimal binning method. The optimal binning for this setting is given when  $\hat{p}_1, \dots, \hat{p}_k$  are all equal to  $1/k$  but achievement of this objective may not always be possible. In addition, a sufficient number of bins must be used to ensure that the process is identifiable but too many bins leads to problems associated with estimating small proportions.

## 3. SIMEST FOR A SINGLE REALIZATION

Again let  $\{N(t), t \geq 0\}$  denote the stochastic process of interest but instead of observing  $n$  i.i.d values of  $\{N(t)\}$  we observe the process at  $n$  different time points, i.e.  $N(t_1), \dots, N(t_n)$ . If one can simulate the process  $\{N(t)\}$ , in theory the SIMEST estimator can be obtained. As an example of the use of SIMEST in this setting consider a general birth and death process.

### 3.1 Simulation of Birth and Death Processes

Consider the counting process  $N(t)$  with parameters  $\lambda_n$  and  $\mu_n$  which satisfies the following axioms:

- i)  $P(N(t + \Delta t) = n + 1 | N(t) = n) = \lambda_n \Delta t + o(\Delta t)$
- ii)  $P(N(t + \Delta t) = n - 1 | N(t) = n) = \mu_n \Delta t + o(\Delta t)$
- iii) The probability of more than one event in  $(t, t + \Delta t] = o(\Delta t)$ .

From the above axioms it is simple to derive the distribution of the time of the next arrival,  $F_B(t)$  and the distribution of the time of the next exit from the system,  $F_D(t)$  so that

$$F_B(t) = 1 - P\{0 \text{ births in } (t, t + \Delta t]\} = 1 - e^{-\lambda_n t}$$

and

$$F_D(t) = 1 - P\{0 \text{ deaths in } (t, t + \Delta t]\} = 1 - e^{-\mu_n t}.$$

Using the inverse c.d.f. transformation we obtain obtain the time until the next birth or death in our process from

$$t_B = \frac{\log(\lambda_n)}{U_1} \quad \text{or} \quad t_D = \frac{\log(\mu_n)}{U_2} \quad (3.1),$$

where  $U_1$  and  $U_2$  represent independent random variables from the uniform distribution defined over the unit interval. To simulate the process  $N(t)$  we use the following algorithm.

#### ALGORITHM TO SIMULATE A BIRTH AND DEATH PROCESS

- 1) Simulate  $U_1$  and  $U_2$  from the uniform(0,1) distribution.
- 2) Compute  $t_D$  and  $t_B$  from (3.1).
- 3) Set  $t = t + \min(t_B, t_D)$ .
- 4) If  $t_D < t_B$  then  $N(t) = N(t) - 1$  else  $N(t) = N(t) + 1$ .
- 5) If  $t < \max t$  and if  $N(t) > 0$  go to 1, otherwise stop.

If there is a maximum population size, say  $N$ , the above algorithm is modified as follows:

#### ALGORITHM TO SIMULATE A BIRTH AND DEATH PROCESS WHEN THE MAXIMUM POPULATION SIZE IS $N$

- 1) Simulate  $U_1$  and  $U_2$  from the uniform(0,1) distribution.
- 2) If  $N(t) = N$  then set  $t_B = \max t$  and compute  $t_D$  from (3.1).  
Otherwise, compute  $t_D$  and  $t_B$  from (3.1).
- 3) Set  $t = t + \min(t_B, t_D)$ .
- 4) If  $t_D < t_B$  then  $N(t) = N(t) - 1$  else  $N(t) = N(t) + 1$ .
- 5) If  $t < \max t$  and if  $N(t) > 0$  go to 1, otherwise stop.

Now that we have established a method for simulating a general birth and death process we again need to consider the question, how should the simulated path and observed series be compared?

### 3.2 Choice of $S_n(\theta)$ .

A reasonable way to proceed is to extend the goodness of fit function discussed in §2.2 to this setting. Let us consider binning the "time" axis into  $k$  bins, namely  $(0, t_1], \dots, (t_{k-1}, t_k]$ . Let  $\hat{n}_1, \dots, \hat{n}_k$  denote the observed value of  $\{N(t)\}$  at the right endpoint of each bin. Let  $\bar{n}_1(\theta), \dots, \bar{n}_k(\theta)$  denote the average value of the  $m$  simulated realizations at the respective times. The goodness of fit function will then be given by

$$S_n(\theta) = \sum_{j=1}^k \frac{(\bar{n}_j(\theta) - \hat{n}_j)^2}{\bar{n}_j(\theta)}. \quad (3.2)$$

In many scenarios a natural binning arises, for example when a birth and death process is observed on a monthly or yearly basis. An important question to consider is whether the above method leads to identifiability problems or can the birth rates and death rates be estimated separately when only the total count is observed?

If both the number of births and the number of deaths are observed a better criterion function would be a weighted average of proximity measures for both curves. More explicitly, let  $\hat{n}_{b1}, \dots, \hat{n}_{bk}$  and  $\hat{n}_{d1}, \dots, \hat{n}_{dk}$  denote the observed number of births and deaths, respectively, of the process  $\{N(t)\}$  at time points  $t_1, \dots, t_k$  and  $\bar{n}_{b1}(\theta), \dots, \bar{n}_{bk}(\theta)$  and  $\bar{n}_{d1}(\theta), \dots, \bar{n}_{dk}(\theta)$  denote the average births and deaths of the  $m$  simulated series. A reasonable comparison between the observed and simulated series would be

$$S_n(\theta) = w \sum_{j=1}^k \frac{(\bar{n}_{bj}(\theta) - \hat{n}_{bj})^2}{\bar{n}_{bj}(\theta)} + (1 - w) \sum_{j=1}^k \frac{(\bar{n}_{dj}(\theta) - \hat{n}_{dj})^2}{\bar{n}_{dj}(\theta)} \quad (3.3)$$

where  $w$  is some appropriately chosen weight function (e.g. the ratio of the total observed births to the sum of the total births and total deaths). By separating births and deaths we avoid any cancelling effect, thereby facilitating the estimation process. Certainly in this situation one can estimate the individual birth and death rates using the SIMEST methodology.

#### 4. THE STATISTICAL PROPERTIES OF THE SIMEST ESTIMATOR

##### 4.1 Strong Consistency of $\hat{\theta}$ .

When the objective is to estimate parameters from  $n$  independent and identically distributed observations from the process  $\{N(t), t \geq 0\}$  the SIMEST procedure leads to strongly consistent estimators of the parameters. Recall, that in this setting the number of occurrences in bins  $1, \dots, k$  follows a multinomial distribution with parameters  $n, p_1(\theta), \dots, p_k(\theta)$  where  $\{N(t)\}$  determines the value of  $p_1(\theta), \dots, p_k(\theta)$ .

*Theorem.* Let  $\hat{p}_j$  denote the observed proportion of  $n$  i.i.d. stochastic processes whose outcome is within  $(t_{j-1}, t_j]$  for  $j = 1, \dots, k$ . Let  $\tilde{p}_j(\theta)$  denote the proportion of the  $m$  simulated processes falling into bin  $j$ . If  $\tilde{p}_j(\theta) \xrightarrow{a.s.} p_j(\theta)$  as  $m \rightarrow \infty$  for  $j = 1, \dots, k$  and all  $\theta \in \Theta$  then  $\hat{\theta} \xrightarrow{a.s.} \theta_0$  as  $m, n \rightarrow \infty$  where  $\theta_0$  denotes the true parameter value,  $\hat{\theta}_n$  is the  $\inf_{\theta \in \Theta} S_n(\theta)$  and  $S_n(\theta)$  is given by (2.1).

*Proof.* For fixed  $n$ , since  $\tilde{p}_j(\theta) \xrightarrow{a.s.} p_j(\theta)$  as  $m \rightarrow \infty$  for all  $\theta \in \Theta$ ,  $S_n(\theta)$  is a continuous function of  $(\tilde{p}_1(\theta), \dots, \tilde{p}_k(\theta))$  and  $0 < p_j(\theta) < 1$  for  $j = 1, \dots, k$

$$S_n(\theta) - S_n^*(\theta) \xrightarrow{a.s.} 0, \quad \text{as } m \rightarrow \infty$$

where

$$S_n^*(\theta) = \sum_{i=1}^k \frac{(p_i(\theta) - \hat{p}_i)^2}{p_i(\theta)}.$$

Now, suppose  $\hat{\theta}_n$  does not converge with  $P_{\theta_0}$  probability 1 to  $\theta_0$ . Then there exists  $\delta > 0$  such that

$$P_{\theta_0}\{\lim_n \|\theta - \theta_0\| \geq \delta\} > 0$$

which implies

$$P_{\theta_0}\{\lim_n \inf_{\|\theta - \theta_0\| \geq \delta} \{S_n^*(\theta) - S_n^*(\theta_0)\} \leq 0\} > 0. \quad (4.1)$$

But by the SLLN  $\hat{p}_i \xrightarrow{a.s.} p_i(\theta_0)$  as  $n \rightarrow \infty$ , therefore  $S_n^*(\theta_0) \xrightarrow{a.s.} 0$  as  $n \rightarrow \infty$  which contradicts (4.1). Hence,  $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$  as  $m, n \rightarrow \infty$ .

The condition requiring that the simulated bin probabilities converge almost surely to the true bin probabilities is addressed by Thompson, *et. al.* (1987); this condition is equivalent to the condition that the process is  $k$ -identifiable.

##### 4.2 Confidence Intervals and Variance Estimates.

One method originally posed by Thompson, *et. al.* (1987), of obtaining confidence intervals for  $\theta$  is to fit a quadratic function to simulated values of the criterion function at locally optimal design points and argue that  $\hat{\theta}$  follows a multivariate normal distribution. Additionally, one can further exploit the simulation feature of the proposed methodology. First, simulate the distribution of  $S_n(\hat{\theta})$  and obtain the upper 95<sup>th</sup> percentile,  $P_{95}$ , of

this empirical distribution. Then evaluate  $S_n(\theta)$  at every point in  $A$ , a square lattice of  $\hat{\theta}$ . The sample covariance matrix  $\hat{\Sigma}$  of the values contained in the set  $A \cap B$  where  $B = \{\theta : S_n(\theta) \leq P_{95}\}$  provides a reasonable estimate of the variation of our estimators. The required sampling distribution can then be approximated by a multivariate normal distribution with mean  $\hat{\theta}$  and covariance  $\hat{\Sigma}$ .

Also of interest in the application of stochastic processes is estimation of the mean path. There are two sources of variation in our SIMEST procedure, namely estimation of the parameter values and the stochastic nature of the process itself. We incorporate both of these components in our estimation of the mean path by first simulating a parameter value from the previously discussed sampling distribution of  $\theta$  and then simulating a realization from the axioms defining the process. This process is repeated at least 500 times thereby yielding an estimate of the distribution of  $N(t)$  for each time point  $t$ . The estimate of the mean path is simply the average at each point in time of the simulated process. Confidence intervals for the mean path are given by the lower 2.5<sup>th</sup> and upper 97.5<sup>th</sup> percentile at each time point with adjustments made for the empirical nature of the limits.

## 5. IMPLEMENTATION OF THE SIMEST PROCEDURE.

In this section, we demonstrate the application of the SIMEST procedure to several successively more complex stochastic processes. We also discuss implementation details. We first apply the SIMEST procedure to the simplest stochastic process, i.e. the homogeneous Poisson process. For two or more bins, the conditions of Theorem 1 are satisfied so that the SIMEST estimator of the Poisson rate does converge almost surely to the true rate. For 1,000 simulated series from a homogeneous Poisson process with rate one, we estimated the rate using the SIMEST procedure with the Nelder-Mead optimization method. The mean and standard deviation for this simulation are .955 and .077, respectively. The results of this simulation would indicate a bias in the procedure, however, as we show later the observed bias is not due to the formulation of the SIMEST procedure but rather with the choice of optimization methods.

A key advantage of this procedure is the ease with which it is parallelized. Much of our work was performed on a Levco parallel processor with seven active nodes. In the simulation studies which follow, each simulated series is the average of 490 individual series (corresponding to seven nodes times seven runs). One crucial step in the SIMEST procedure is optimization of the function  $S(\theta)$  over the parameter space. Simulation studies were performed using both the Nelder-Mead optimization algorithm (N-M) and a new algorithm proposed by Lawera and Thompson (1992) (L-T). Using Nelder-Mead the simulations result in biased and highly correlated estimates of the model parameters whereas much better results are obtained via the algorithm of Lawera and Thompson.

### 5.1 Application to Birth and Death Processes

To illustrate that indeed the SIMEST estimation procedure can recover the true parameters of a stochastic process, we simulated the mean path over 20 bins for the four models given in Table 1 (again based on 490 realizations) and applied the SIMEST algorithm with proximity measure (3.3) and N-M optimization to the mean path. As expected, we obtained very good results whenever the maximum population size  $N$  was known. Model IV of Table 1 depicts the situation when  $N$  is unknown and is also estimated from the given data. As suggested by Tapiero (1975) we estimate all the parameters but restrict  $N$  to mutually exclusive intervals. The final estimates are then chosen from the set of estimates within each interval, again so that the criterion function  $S_n(\theta)$  is minimized. Of course,



much more work should be conducted on estimation of  $N$  before one can reliably use this as an option in data analysis. The results of the simulations are summarized in Table 2.

Consider now the empirical study of the statistical properties of the SIMEST estimator for birth and death processes. Table 3 list the results of the different studies for each model whereas Figure 1 illustrates graphically the important points. For Model I the SIMEST procedure with N-M optimization was applied to 1,000 simulated realizations for the true model. The estimates for this model were unbiased and closely follow a multivariate normal distribution. The main problem is the strong correlation between the two parameter estimates (.88). This is due to the choice of optimization routines not to the fundamental concept of SIMEST. For Model II with parameters  $\lambda = 0.01$  and  $\mu = 0.03$  (different from those listed in Table 1) 100 replications of the SIMEST estimator using N-M optimization were obtained. When broad starting values are used the estimator is severely biased downwards as can be seen in Figure 1. Attempts to reduce this bias leads to strong correlation between the two estimates. The problems of bias and correlation are due to the choice of optimization routines and not to the conceptual underpinnings of the SIMEST methodology. To illustrate this fact, two additional simulation studies were performed using the newly developed optimization routine of Lawera and Thompson (1992). Unbiased and uncorrelated estimates were obtained for Models II and III. Figure 1 provides a scatterplot of 250 estimates from Model II and histograms of 250 estimates of each parameter in Model III.

## 6. CONCLUSIONS

SIMEST provides the researcher with the ability to implement the appropriate stochastic model without the arduous task of solving complicated differential or difference equations. For multiple independent observations from one stochastic process, SIMEST leads to consistent estimators of the process. SIMEST easily recovered the correct parameters when the mean path of the complicated birth and death process (with  $N$  fixed) was input. However, for varying  $N$  the estimates obtained from SIMEST are suspect; more work in this area is necessary. Repeated applications of SIMEST to single realizations illustrated that using the Lawera-Thompson optimization routine the SIMEST estimator possesses desired statistical properties (unbiased and small variance).

Table 1. Birth and Death Processes Considered

MODEL	$\mu_n$	$\lambda_n$	PARAMETER VALUES			
			$\mu$	$\lambda_1$	$\lambda_2$	$N$
I	$n\mu$	$n\lambda_1$	.25	.35		
II	$n\mu$	$(N - n)n\lambda_1$	.05	.04		800
IIb	$n\mu$	$(N - n)n\lambda_1$	.10	.03		800
III	$n\mu$	$(N - n)n/N\lambda_1 + (N - n)\lambda_2$	.05	.10	.05	800
IV (N ESTIMATED)	$n\mu$	$(N - n)n/N\lambda_1$	.05	.04		800

Table 2. *SIMEST Estimates of the Mean Path*

MODEL	PARAMETER VALUE	SIMEST ESTIMATE	ESTIMATED STANDARD DEVIATION
I	.35	.3546	.008811
	.25	.2565	.011156
II	.04	.0399	.000555
	.05	.0495	.001069
III	.05	.0496	.000578
	.10	.1005	.000577
	.05	.0491	.000629
IV	.04	.04068	.001160
	.05	.05020	.001137
	800	806.7	19.644

Table 2. *Replicated SIMEST Estimates From Single Realizations*

MODEL	OPTIMIZATION METHOD	NUMBER OF REPLICATIONS	PARAMETER VALUE	SIMEST ESTIMATES MEAN	STD. DEV.
I	N-M	1,000	.35	.3371	.04362
			.25	.2398	.04295
IIb	N-M	100	.10	.08480	.01473
			.03	.02354	.00774
II	L-T	250	.04	.04022	.004278
			.05	.05008	.003765
III	L-T	250	.05	.04935	.01289
			.10	.09897	.00835
			.05	.04992	.00643

Figure 1. Simulation Results for Models I, II and III

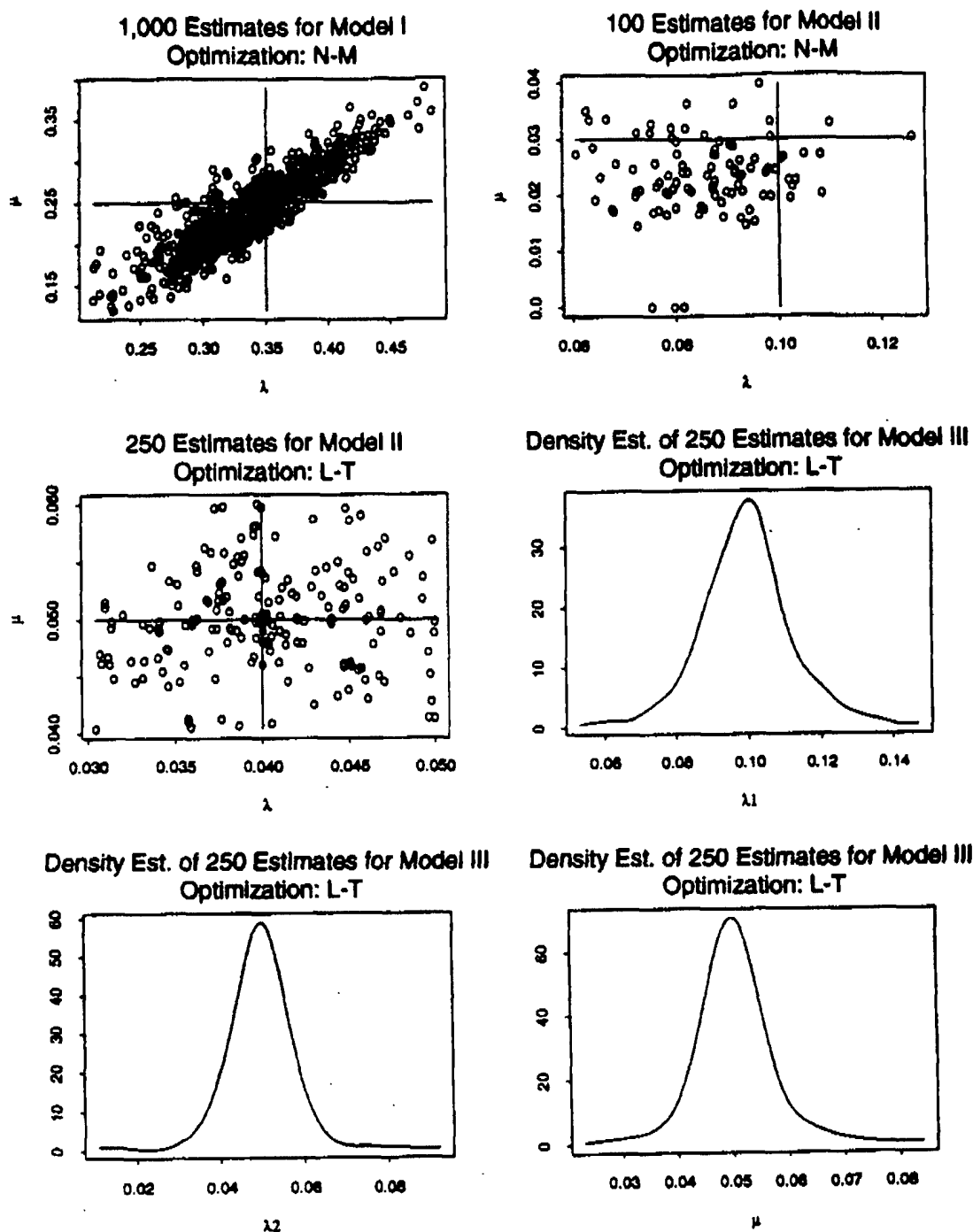


Figure 1: The above figure represents the results of the simulation studies performed on the SIMEST procedure. The title of each graph gives the model number corresponding to the models of Table 1. The first three graphs are scatterplots of the estimates of the two parameters for the given model. Axes are drawn at the true parameter values. The last three plots are kernel density estimates for the estimates of parameters of Model III, namely, .1, .05, and .05.

#### REFERENCES

- Bartholomew, D.J. (1982). Stochastic Models for Social Processes, John Wiley and Sons, NY., p.311-312.
- Bartlett, M.S. (1960). Stochastic Population Models in Ecology and Epidemiology, John Wiley and Sons, Inc., NY, p.55-56.
- Bridges, Eileen, Katherine B. Ensor, and James R. Thompson (1992). "Marketplace Competition in the Personal Computer Industry," *To appear in Decision Sciences*.
- Eliashberg, Jehoshua, and Rabikar Chatterjee (1986). "Stochastic Issues in Innovation Diffusion Models," Innovation Diffusion Models of New Product Acceptance, Ballinger, eds. Vijay Mahajan and Yoram Wind, Publishing Company, Cambridge, MA, 151-199.
- Lawerà, M. and Thompson, J. R. (1992). "A Parallelized, Simulation Based Algorithm for Parameter Estimation," to appear in the *Proceedings of the Thirty-Seventh Conference on the Design of Experiments in Army Research, Development, Testing*.
- Nelder, J. A. and Mead, R. (1965). "A simplex method for function minimization." *Computational Journal*, 7, 308-313.
- Tapiero, C.S. (1975). "On-Line and Adaptive Optimum Advertising Control by a Diffusion Approximation," *Operations Research*, 23, 890-907.
- Thompson, J. R., Brown, B. W., and Atkinson, E. N. (1987). "SIMEST An Algorithm for Simulation-Based Estimation of Parameters Characterizing a Stochastic Process," Cancer Modeling, eds. J. R. Thompson and B. W. Brown. Marcel Dekker, p. 387-415.
- Thompson, J. R. and Tapia, R. A. (1990). Nonparametric Function Estimation, Modeling, and Simulation, SIAM, Philadelphia.

# **The Thirty Eighth Conference on the Design of Experiments in Army Research, Development and Testing**

**RAND, Santa Monica, California**

**28 - 30 October 1992**

## **Attendance List**

<b><u>NAME</u></b>	<b><u>AFFILIATION</u></b>
John Adams	RAND
Gerald R. Andersen	US Army Research Office
William E. Baker	US Army Research Laboratory
Russell R. Barton	Penn State University
Carl B. Bates	US Army CAA
MAJ Kevin M. Beam	US Army TRADOC, RAND
Robert M. Bell	RAND
Jeffrey P. Benedict	NSA
L. Mark Berliner	Ohio State University
Barnard H. Bissinger	US Navy Ships Parts Control Center
Barry A. Bodt	US Army Research Office
Ann M. E. Brodeen	US Army Research Laboratory
Melvin Brown	US Army Research Office
Marshall N. Brunden	The UPJOHN Company
Marion R. Bryson	US Army TEXCOM
Robert J. Burge	Walter Reed Army Institute of Research
J. Steve Caruso	US Army Management Engineering College
Aivars Celmins	US Army Research Laboratory
COL C. Terry Chase	US Army SLA
James Chrissis	US Air Force Institute of Technology
W. J. Conover	Texas Tech University
Noel A. C. Cressie	Iowa State University
Terrence M. Cronin	US Army CECOM
David F. Cruess	USUHS Medical School
Francis E. Dressel	US Army Research Office
Naihua Duan	RAND
Eugene F. Dutoit	US Army Infantry Warfighting Center
Marc N. Elliot	Rice University
Samuel Frost	US Army MSAA
Lionel Galway	RAND
Donald P. Gaver	Naval Postgraduate School
LCDR Robert J. Gregg	US Navy, RAND
Joe Harmon	US Army TEXCOM

**The Thirty Eighth Conference on the Design of Experiments  
in Army Research, Development and Testing**

**RAND, Santa Monica, California**

**28 - 30 October 1992**

**Attendance List (continued)**

<b><u>NAME</u></b>	<b><u>AFFILIATION</u></b>
Bernard Harris	University of Wisconsin
James S. Hodges	RAND
Charlie Holman	US Army OEC
William Jackson	US Army TACOM
Ronald Leon Johnson	US Army Belvoir RD&E Center
Daniel Todd Jones	US Army OEC
W. D. Kaigh	University of Texas at El Paso
Martin Lawera	Rice University
Thomas W. Lucas	RAND
James R. Maar	NSA
Etan Markowitz	Private Consultant
Daniel McCaffrey	RAND
Sally C. Morton	RAND
Donald Neal	US Army MTL
Emanuel Parzen	Texas A&M University
Michael P. Prather	US Army TECOM
Daniel A. Relles	RAND
Jorma J. Rissanen	IBM Research Center
John E. Rolph	RAND
Ernest G. Scherb	Natick RD&E Center
David W. Scott	Rice University
Jayaram Sethuraman	Florida State University
Nozer D. Singpurwalla	George Washington University
Douglas B. Tang	Walter Reed Army Institute of Research
Malcolm S. Taylor	US Army Research Laboratory
Delores Testerman	US Army TEXCOM
Jerry Thomas	US Army Research Laboratory
James R. Thompson	Rice University
Henry B. Tingey	University of Delaware .
Mark G. Vangel	US Army MTL
David W. Webb	US Army Research Laboratory
Franklin E. Womack	US Army CAA